DOCUMENT RESUME

ED 282 508 HE 020 424

TITLE	The Statewide Master Plan, State of Washington Higher Education Coordinating Board. Issue Area: Quality. Question No. 4, Performance Evaluation.
INSTITUTION	Washington State Higher Education Coordinating Board, Olympia.
PUB DATE	Jan 87
NOTE	104p.; Paper collected as part of the American Association for Higher Education Assessment Forum.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Academic Achievement; College Graduates; *College Students; *Educational Assessment; Educational Quality; Educational Testing; Education Work Relationship; Employers; Higher Education; Job Satisfaction; Majors (Students); Master Plans; Mastery Tests; *Outcomes of Education; *Statewide Planning; Student Development; *Student Evaluation
IDENTIFIERS	*AAHE Assessment Forum; Value Added; *Washington

ABSTRACT

Options for evaluating the performance of Washington state colleges based on student outcomes are discussed. Considered in "Question Number 4" on the agenda of the Washington State Higher Educaton Coordinating Board: how should the performance of higher education institutions be evaluated and by whom? Of concern is what student outcomes to measure, whether student progress should be influenced by the measure, and whether the evaluation should be designed at the state or college level. Three ways of assessing student performance provide useful data about college performance: assessment of student achievement, graduate placement rates, and alumni satisfaction surveys that indicate cognitive and affective student outcomes. Advantages/disadvantages of the following assessment options are discussed: sophomores and seniors should be assessed to determine communication and computation skills; seniors should be assessed to determine mastery of their major; graduates should be assessed to determine subject matter proficiency; and colleges should summarize the job placement of graduates and their satisfaction with their education, as well as employers' satisfaction with graduates. Attached is a 64-page report prepared for the Washington Higher Education Coordinating Board, and entitled "The Evaluation of Institutions of Higher Education in the State of Washington." The report includes Appendixes A through G, on: Testing (Gerald M. Gillmore); Value Added Testing (Robert M. Thorndike); Assessment of Alumni (Fred E. Fiedler and Douglas J. Allmon); Sample Questions; Expense Considerations; and Reviewer Comments. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED282508

THE STATEWIDE MASTER PLAN

Issue Area: Quality

Question #4 - Performance Evaluation

The Higher Education Coordinating Board invites public testimony on this question at its February 17, 1987, meeting at the Vance Airport Inn (SeaTac).

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Washington H.E. Coordinating Board

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Higher Education Coordinating Board
State of Washington

January, 1987

HE 020 424

BEST COPY AVAILABLE

One Dupont Circle
Suite 600
Washington, D.C. 20036
202/293-6440

AMERICAN ASSOCIATION
FOR HIGHER EDUCATION

Board of Directors

Chair
Joseph F. Kauffman
University of Wisconsin
Madison

Chair-Elect
Adele S. Simmons
Hampshire College

Vice Chair
Reatha Clark King
Metropolitan State University

Past Chair
Harriet W. Sheridan
Brown University

Carlos H. Arce
NuStats, Inc.

Estela M. Bensimon
Teachers College
Columbia University

Anne L. Bryant
American Association of
University Women

Donald L. Fruehling
McGraw-Hill, Inc.

Ellen V. Futter
Barnard College

Jerry G. Gaff
Hamline University

Zelda F. Gamson
University of Michigan

Stephen R. Graubard
Daedalus

Joseph Katz
State University of New York
at Stony Brook

Arthur E. Levine
Bradford College

Frank Newman
Education Commission
of the States

Alan Pifer
Carnegie Corporation
of New York

W. Ann Reynolds
The California State
University

Piedad F. Robertson
Miami-Dade Community
College

D. Wayne Silby
Groupware Systems

P. Michael Timpane
Teachers College
Columbia University

President
Russell Edgerton

The AAHE ASSESSMENT FORUM is a three-year project supported by the Fund for the Improvement of Postsecondary Education. It entails three distinct but overlapping activities:

--an annual conference
(the first scheduled for June 14-17, 1987, in Denver)

--commissioned papers
(focused on implementation and other timely assessment concerns; available through the Forum for a small fee)

--information services
(including consultation, referrals, a national directory, and more)

This paper is part of an on-going assessment collection maintained by the Forum. We are pleased to make it more widely available through the ERIC system.

For further information about ASSESSMENT FORUM activities, contact Patricia Hutchings, Director, AAHE ASSESSMENT FORUM, One Dupont Circle, Suite 600, Washington, DC 20036

CHARLES T. COLLINS
Chair

A. ROBERT THOENY
Executive Director

STATE OF WASHINGTON
HIGHER EDUCATION COORDINATING BOARD
908 East Fifth Avenue, EW-11 • Olympia, Washington 98504 • (206) 753-2210 • (SCAN) 234-2210

January 9, 1987

MEMORANDUM #87-2

TO: Members of the Higher Education Coordinating Board

FROM: A. Robert Thoeny, Executive Director

SUBJECT: Critical Question #4 - Performance Evaluation

Attached is a paper presenting alternatives in response to Critical Question #4. These materials are presented for Board discussion at the January meeting. Interested parties will be provided copies of this paper and are invited to provide reactions to Board staff within the next two months. Those wishing to address the Board relative to the topic of this paper are invited to do so at the Board's meeting scheduled for February 17, 1987, at the Vance Airport Inn at SeaTac. Prior notification to the Board staff would be appreciated. Board decisions in answer to Critical Question #4 are scheduled for March 17, 1987.

SUMMARY OF OPTIONS FOR QUESTION #4: *How should the performance of higher education institutions be evaluated and by whom?*

This paper focuses on an important purpose of colleges and universities, the education of students. The paper assumes that higher education institutions should be evaluated in terms of the performance of their graduates. Other important purposes of higher education institutions, such as research and public service, would also be proper subjects for evaluating institutional performance, but these are not the focus of this analysis.

This critical question addresses what student outcomes should be measured, if student progress should be influenced by the measure, and whether the evaluation program should be designed at the local institutional level or developed on a statewide basis. Descriptive information about each of the option sets is included in the text of the paper which follows this summary.

What Student Performance Should Be Evaluated?

Option-Set #1: Assessment of Student Achievement

Option #1a: Sophomores should be assessed to determine their communication and computation proficiencies.

Option #1b: Seniors should be assessed to determine their communication and computation proficiencies.

Option #1c: Seniors should be assessed to determine mastery of their major field.

Option #1d: Graduate students should be assessed to determine their proficiency in their subject matter.

Option-Set #2: Graduate Job Placement

Option #2a: Community colleges should summarize the job placements of their vocational and technical program graduates.

Option #2b: Four-year institutions should summarize the job placements of their four-year graduates.

Option #2c: Four-year institutions should summarize the job placements of students from their graduate programs.

Option-Set #3: Alumni/Employer Satisfaction

Option #3a: Community colleges and four-year institutions should summarize graduates' satisfaction with the quality of their college experience.

Option #3b: Community colleges and four-year institutions should summarize employers' satisfaction with employee's college preparation.

How Should a Performance Evaluation System Affect Students?

Option #1: If a statewide program of assessment is adopted, institutions should be evaluated on student data which are summarized.

Individual student progress should not be directly affected by the results of any assessment.

Option #2: If a statewide program of assessment is adopted, individual students should not be permitted to enter the next educational level until they demonstrate the required proficiencies.

Who Should Be Responsible for Performance Evaluation?

Option #1: Each public institution should be responsible for developing and implementing a comprehensive performance evaluation plan.

Option #2: A statewide performance evaluation program should be established for public higher education institutions which allows comparisons to be made among these institutions.

Option #3: A statewide performance evaluation program should be established. However, each institution would be able to add to the program to meet local needs.

INTRODUCTION

With limited state resources and increasing demands on all public agencies, questions have been raised about the performance of higher education institutions. Educators would agree that the postsecondary educational experience should positively impact a student's knowledge base as well as the student's attitudes and skills. And, yet, how does an institution demonstrate that a student has gained the prescribed knowledge and skills upon completion of a program?

Often, the faculty of higher education institutions, legislators, and the public are uncertain if graduates have learned what they need to know in order to contribute at a high level once they are in the work force. Are graduates leaving educational institutions with the intended skills and knowledge base? Have four-year graduates developed sophisticated thought processes which include the abilities to analyze and synthesize data? Do they have a broad understanding of western civilization and its impact on today's events as well as an appreciation for other cultures? In addition to a breadth of knowledge, do graduates possess a depth of understanding about a particular subject matter area? Are these individuals moving into the work place and filling the intended positions in desirable ways? Are students in community college academic programs able to transfer to four-year institutions and be competitive, and are vocational/technical students able to acquire jobs and perform in the work setting?

Since the beginning of this decade, a great deal of attention has been focused on the outcomes of postsecondary education. Institutional evaluation historically has been an important dimension of higher education. However, in the past it centered on teaching and learning processes rather than the

outcomes of these processes. Accreditation teams evaluated an institution's faculty, services, and resources. Peer review panels assessed a department's faculty: the number of publications produced, the research performed, and the credentials and experience of the faculty members.

In outcomes oriented evaluation, institutions are measured against their ability to accomplish their stated goals. This type of evaluation provides holistic information on a program's worth to students, faculty, and decision makers. The intent of this newer focus is to insure that an institution is achieving its goals and, if not, to make adjustments in order to produce the intended results. Through this type of outcomes evaluation effort, overall institutional performance is enhanced which results in an improved teaching and learning environment.

WHAT IS OUTCOMES ASSESSMENT?

The outcomes of an educational experience include a student's cognitive development during the college experience as well as the graduate's career achievement upon completion of the academic program. Student achievement, in terms of the gains made in one's knowledge base, are cognitive outcomes. Affective outcomes include the changes in a student's attitudes and values which occur as a result of the educational process. Occupational and career development, or success in the job market, may also be directly related to a student's educational accomplishments and a measure of the outcomes of a particular educational program.

The following terms are often used, sometimes interchangeably, in discussing performance evaluation. Definitions are provided here to clarify their use and to confirm the

focus on outcome evaluation measures for state policy decisions.

Assessment — a mechanism for prediction and evaluation. An effort which is designed to demonstrate the worth or value of something. The expert judgment involved is based on careful observation. Assessment produces descriptive information about what an individual can do with a knowledge base.

Testing — provides quantitative information elicited through a measurement process about what and how much a person knows.

Evaluation — focuses on elements which may be combined to form summary conclusions about students as a group. Emphasis is placed on decision making about the direction of a course, program, or curriculum.

Formative — evaluation covering a defined segment of a program and conducted as the program unfolds. This type of evaluation is used immediately to modify, delete, or revise the program components.

Summative — evaluation of the success or failure of the whole program in achieving the defined goals. This type of evaluation generally is comprehensive in nature and reviews the entire program (a summative evaluation may take into consideration information gained from formative evaluations).

Value added — an economic term adopted by educators which refers to the value that has been added as a result of academic achievement. Value-added assessment implies the administration of a pretest and posttest.

Testing, assessment, and evaluation occur at all levels within higher education institutions. One traditional form of evaluation is provided by faculty members to students (in the form of grades) at the end of a course. The grade is generally based upon a series of tests and possibly other forms of assessment measures. Course evaluations tend to focus on the students and faculty as well as the content of the particular course under evaluation. Institutions typically engage in program evaluations.

Although these evaluation endeavors provide students, faculty, and campus administrators with useful data about the institution, the information does not provide a collective picture of an institution's ability to achieve results. Because the information is difficult to summarize, conclusions about the institution's success in reaching stated goals cannot be drawn.

In order to effectively measure the educational experience, there must be a clear understanding of the institution's role and mission. Additionally, to adequately gauge the impact of the experience, institutions need to define when they expect the impact to be realized. Will the outcomes result immediately or will they be exhibited after a graduate leaves the institution and embarks upon a career? Does the institution have specific expectations of graduates or are its expectations more general?

The next step is to identify the best measure to use in determining whether the expected outcome has been attained and at what point in the student experience to use that measure. Formative evaluation data provide assistance to the institution in making necessary internal adjustments to the student's program before the student completes the program. Summative information, such as graduate placement records and/or graduate satisfaction indices, provides a means of evaluating the overall effectiveness of the program.

WHEN IS ASSESSMENT APPROPRIATE?

Cognitive development is tested primarily at three important stages in a student's experience. Assessment programs are implemented at these points for very different reasons. The first stage of assessment often occurs when a student enters an institution and focuses on demonstration of competencies in basic skill areas such as reading, writing, mathematics, and oral communication skills. Results of the assessment determine course placement, remediation needs, and baseline data for value-added analysis of the instructional program.

The next logical assessment of cognitive development occurs at the end of the sophomore year. Rising junior examinations establish whether or not a student is prepared to begin upper-division course work. Assessment at this point also indicates if community college students have mastered general education requirements prior to transfer to four-year institutions. An additional benefit of testing at this point is to determine which program or major a student might successfully pursue during the next two years. A program focusing on value-added instructional activities would also benefit from analysis at this point.

Before the baccalaureate degree is conferred, the institution may also decide to examine a student's proficiency generally or specifically. The focus of assessment at this point may be on the development of higher order thinking and reasoning skills and the application of knowledge. Institutions may also pursue value-added analysis at this point in order to determine the student gain in knowledge over the four-year period.

Affective or student opinion assessment may occur throughout a student's experience. In addition, potential students may be surveyed. Their perceptions may be extremely useful in the design or redesign of recruiting strategies. Moreover, alumni may be surveyed to provide information from their perspective as recipients of educational services. Their views regarding the usefulness of a program as it applies to occupational and career demands is critical information. These opinion surveys on the attitudes of potential, current, and previous students are invaluable instruments in providing an institution with a broad range of data about its programs and services.

ASSESSMENT TRENDS

Since many states are interested in "certifying" the results of an educational experience, student testing has exhibited a positive trend. Several states have implemented testing programs at the point of admission to detect the preparedness of students to perform at the college level. The results of the assessment, along with other data about the students, are used to advise students about the appropriate course choices.

Assessment of this type is currently the practice for the majority of community colleges in this state. Upon entry,

students are assessed and advised on course selection and, if needed, advised to enroll in developmental studies if their competency levels in basic skills are not sufficient for college-level work. An additional example of this type of program is the one initiated in New Jersey. A statewide assessment program was implemented there in 1977 called the New Jersey Basic Skills Placement Test. Under this program students are tested in elementary algebra, computation, reading comprehension, and sentence sense. Based on the results, students are advised on appropriate course selections. The California State University system uses a similar exam as a diagnostic tool for course placement.

Some states are requiring students to test to determine their readiness to enter the next educational level. Florida's College Level Academic Skills Test (CLAST) is required of all sophomore students prior to receiving an associates degree or being granted admission to the junior level. Communication and computation skills (such as reading, writing, speaking, listening, algorithms, concepts, generalizations, and problem solving) are measured. Miami-Dade Community College has achieved national recognition for designing the curriculum and support systems necessary to provide students with the knowledge and skills required to be successful under the Florida testing requirements.

Within certain academic fields, testing is often required. For example, some states now require potential students to take an admission test prior to entry into teacher education programs. In this state, teacher education schools require applicants to score a minimum of 80 on the Washington Pre-College Test.

Individual institutions also are developing their own "rising junior" requirements. The University of Massachusetts

at Boston has in place a writing proficiency examination. Undergraduates must demonstrate proficiency before they may enroll in upper division classes. Western Washington University requires all students who have completed 90 credits to take a test which demonstrates their proficiency level in language and writing skills. Although students who do not demonstrate proficiency may continue to take upper-division course work, they must enroll in additional writing courses or auxiliary programs designed to help them develop the necessary skills to be successful. Following successful completion of the proficiency examination, students enroll in an upper-division, "writing-intensive" course.

Some states are requiring students to pass a test prior to graduation. Because of the difficulty in defining what should be tested and the diversity in the core curriculum in American higher education, states involved in assessment at this point have developed different approaches. In Georgia, students are required to pass a "Regents Exam" before they graduate. Of note, students are able to take the test as sophomores and are able to retake the test until they are able to pass it. Another example of a state which requires testing prior to graduation is South Dakota. In that state college majors are required to demonstrate a certain level of performance on tests in their major field as a criterion for successful completion of a program. Some states are enacting laws which require students in certain areas to be tested or "certified". Graduates of baccalaureate programs in certain fields in Mississippi and Missouri must pass a common standardized test as a condition of graduation.

In some cases states have implemented a funding structure which encourages assessment. Tennessee's incentive-funding program rewards institutions in that state which can demonstrate that: (1) a percentage of their programs have under-

gone peer review; (2) a percentage of their programs have administered a comprehensive examination to majors within a five-year period; (3) accomplished both with maximum credit extended if the institution can document that the student experience on the field examination improves over time or if it exceeds that of students in similar programs at comparable institutions; (4) value, in fact, has been added by the general education component of the curriculum as measured by the American College Testing program College Outcomes Measures Project examination; (5) students, alumni, employers, and/or community members have an opportunity to express their opinions about the quality of academic programs and services; and (6) a campus-wide plan has been implemented to improve the instructional program based on the results of the other assessment procedures as well as utilization of other measures and inputs.

USES OF OUTCOME MEASUREMENTS

Institutions currently measuring outcomes report a variety of benefits from the process. At Northeast Missouri State University, the institution most often associated with value-added assessment, three primary goals were delineated: (1) the desire to know everything possible about the student, (2) the importance of demonstrating that the university experience made a positive difference in the student's life, and (3) the interest in demonstrating that graduates of the university were nationally competitive.

Of importance to most institutions who are involved in this process is the focus placed on teaching and learning. In addition, faculty members become more aware of national expectations of university graduates and have a better understanding of how competitive one's own students fare in

the market place. Assessment often becomes the catalyst for curricular change and improved decision making.

Students, faculty, and administrators all realize that modifications to programs and services can be made based upon the findings of the various assessments. Often, the result of a performance evaluation program is that curricula are reshaped, student services are revised or evaluated, more effective strategies are developed to recruit and retain special student populations, programs are evaluated in order to justify the need for additional funding, or a new emphasis is placed on the development of a particular skill.

Because performance has been documented, universities are able to communicate to their various constituencies what they and their students are able to accomplish. Institutions are able to communicate their concern with quality to students and various publics. As a result of outcomes evaluation, aggregate information about an institution is available, providing the legislature with measures for accountability purposes.

Additional benefits occur when institutions are able to improve program articulation with high schools in terms of the entry level skills which are expected of their entering freshmen. Communication with graduate schools and professions is also enhanced, since the expectations for the program are clearly stated and the instruction is designed to accomplish specified results.

POLICY OPTIONS

The option sets for performance evaluation are presented below. As a part of its efforts to gather information on outcomes assessment, the Higher Education Coordinating Board staff contracted with Dr. Fred Fiedler to study the technical

feasibility of implementing a performance evaluation program. The report resulting from this study (enclosed in Attachment) provides information which is pertinent to a number of the options.

It should be noted that performance evaluation efforts take time to design and to implement and are costly. Should the Higher Education Coordinating Board decide to include recommendations for a comprehensive evaluation program in the plan, budgetary support to design, implement, and maintain these programs will be required.

WHAT STUDENT PERFORMANCE SHOULD BE EVALUATED?

Three methods of collecting data about student performance provide useful information about institutional performance. Assessment of student achievement, graduate placement rates, and alumni satisfaction questionnaires indicate cognitive and affective student outcomes. Descriptions of each of these methods are provided below along with a set of possible options.

Assessment of Student Achievement

Most assessment of student performance involves a test of achievement. The test's purpose should determine when the examination is administered and whether all or a representative sample of students are tested. Testing cognitive achievement can take place at different points in a student's experience: at the end of the sophomore year, prior to receipt of the baccalaureate degree, and prior to completion of a graduate program.

Option #1a: Sophomores should be assessed to determine their proficiency levels in communication and computation.

Assessment of communication and computation skills at the end of the sophomore year is an effective method of determining whether all students have developed the required proficiency in basic skills at that point. With students of varying abilities able to gain access to Washington's institutions of higher education, it may be beneficial to provide a standard for communication and computation skills at the end of the first two years.

Advantages

* Students have a clear understanding of the proficiency levels of skills that are required for success in upper-division courses.

* Students who do not demonstrate proficiency may be advised into appropriate classes which will assist them in developing the necessary skills.

* In a state with a large community college system, better articulation between the two-year and four-year institutions may be the result of a sophomore assessment program.

* Students who are not developing the required skill levels find out in time to remedy the situation.

Disadvantages

* An assessment program for all sophomore students in the state will be expensive to implement.

* A statewide assessment program is likely to test only fundamental skills in reading, writing, and computing. Other expected outcomes, such as higher level thinking skills and synthesizing abilities, which are more difficult to assess, probably will not be included in order to keep the program administratively practical and economical.

Option #1b: Seniors should be assessed to determine their communication and computation proficiency levels.

Assessment of seniors prior to graduation would demonstrate if the students have attained the desired proficiency in specific skill areas. With legislatures, governing boards, and the public interested in the outcomes of higher education, institutions would be able to "certify" that their students have developed generally agreed-upon abilities.

Advantages

* Faculty and administrators of baccalaureate institutions would have a data set about the skill levels of seniors. These data could be useful in making curricula decisions at the upper-division as well as the lower-division level.

* Institutions could make use of the full four-year curriculum in developing the desired cognitive abilities of students. Many skills are mastered or enhanced through the upper-division curriculum.

Disadvantages

* Implementation of a statewide senior assessment program would be costly.

* Assessing the basic skill development of seniors reduces the value of a four-year program to the "lowest-common denominator".

* Assessing general education at this point in a student's career is too late to provide remediation of the student's weaknesses.

Option #1c: Seniors should be assessed to determine mastery of their major.

Another appropriate outcomes assessment of seniors which concerns the depth of knowledge deemed important by a particular academic department is the assessment of students in their major field of study. This type of assessment provides faculty with direct and important information about the upper-division curriculum.

Advantages

* Departments would be able to "certify" that their seniors have acquired a stated body of knowledge.

* Faculty and employers could collaborate on defining the important elements students should acquire in a particular major area.

* Students would know how they compared to other students in their major.

Disadvantages

* Deciding on the appropriate assessment instruments for subject matter tests is a time-consuming and costly proposition.

* In the disciplines of the liberal arts and sciences, application of a standardized assessment instrument could lead to a standardized curriculum, a result which would be judged by most educators to be undesirable.

Option #1d: Graduates should be assessed to determine proficiency in their subject matter.

Although most recent attention to outcomes assessment has been focused on undergraduate programs, it may also be useful to evaluate institutional performance by assessing the performance of graduate students. This could be accomplished with least difficulty at the master's level in professional fields, and with greater difficulty at the doctoral level in the liberal arts and sciences disciplines.

Students in graduate programs should be able to demonstrate that they have attained the competencies required for their particular field of study. These competencies should be defined by the faculty members who have developed the curriculum.

Advantages

* Graduate schools would be able to "certify" their students.

* Graduate students would be able to compare themselves with other students in that field of study.

Disadvantages

* Most graduate programs are in essence programs of individualized studies. The evaluation of these

programs by means of a common assessment instrument would be extremely difficult.

* Implementation of a statewide program would be time-consuming and costly.

Graduate Job Placement

Job placement data are excellent indicators of a program's viability. The placement rate not only indicates the strength of a particular program's reputation but also is an indication of the quality of its curriculum. Often, communication between graduates and faculty resulting from this type of summary is useful in making adjustments to the curriculum.

Option #2a: Community colleges should summarize the job placements of completers of their vocational and technical programs.

Follow-up surveys of program completers to determine whether the former student is employed and in which job field could be undertaken. Results would yield a rate of placement in fields for which the student was trained.

Advantages

* Placement rate data are useful in the continuing need for evaluating programs.

* Follow-up provides faculty with an important link to business and industry.

* Information which is gained from follow-up studies may influence curricula decisions.

* Making placement rate information available to potential students would usefully inform their choice of vocation.

Disadvantages

* Compiling placement data is time-consuming and expensive.

* It is often difficult to track graduates over a period of years, if longitudinal data are desired.

Option #2b: Four-year institutions should summarize the job placements of graduates from their baccalaureate programs.

Again, follow-up surveys of graduates could provide a useful measure of institutional success, although the development of a rate of placement in fields for which the student was prepared would be very difficult outside of the professional schools.

Advantages

* Summary data on job placements are effective student recruiting tools for college departments.

* Contacts with students in the job market provide faculty with a good mechanism for evaluating the effectiveness of their programs.

Disadvantages

* Student follow-up is time-consuming and expensive.

* Over time, it may be difficult to maintain good records of students' work history.

Option #2c: Four-year institutions should summarize the job placements of students from their graduate programs.

Since preparation at the graduate level is generally more specific than undergraduate preparation, universities will have less difficulty determining appropriate placement.

Advantages

* Summary data on job placements are effective student recruiting tools for college department.

* Contacts with students in the job market provide faculty with a good mechanism for evaluating program effectiveness.

Disadvantages

* Student follow-up is time-consuming and expensive.

* Over time, it may be difficult to maintain good records of students' work history.

Alumni/Employer Satisfaction

An important source of information about the quality of a college program is the extent to which graduates are satisfied with their experience. An additional source of information about a graduate's experience, is how satisfied a graduate's employer is with the college preparation the student received.

Option #3a: Community colleges and four-year institutions should summarize graduates' satisfaction with the quality of the college experience.

Program graduates may be the best judges of the value-added by their college experience. This measurement goes beyond job placement to capture the non-vocational benefits of having attended college.

Advantages

* Information gained from graduates' questionnaires may result in changes made to the curriculum and improvements in the services offered to students.

* Graduates are able to provide data about their college experiences which is summative in nature. This type of evaluation produces a useful, holistic view of the value of their education.

Disadvantages

* Information from surveys of this nature is expensive to collect, analyze, and report.

* The data which are generated from the survey may not be specific enough to be useful.

Option #3b: Community colleges and four-year institutions should summarize employers' satisfaction with employee's college preparation.

The information gathered by this effort would compliment job placement rate data. The quality and currency of the curriculum could be evaluated.

Advantages

* Information gained from employee questionnaires may result in improved understanding of the desired outcomes of the program. As a result, changes may be made to the curriculum.

* Enhanced communication may result between faculty members and employers which may help placement rates.

Disadvantages

* Alumni surveys are time-consuming and expensive.

* The results of the survey may not be specific enough to be useful.

HOW SHOULD A PERFORMANCE EVALUATION SYSTEM AFFECT STUDENTS?

A very important consideration in the design of an assessment system is the matter of its applicability. Should the system be used to pass judgment upon each student's qualifications for receiving a credential or should student scores be used only as a means of evaluating institutional performance or should the assessment system do both of these. Important consequences follow. For example, if individual student evaluation is not a purpose, institutions can be evaluated through assessment of a sample of their students, but if assessment is part of the student's demonstrated qualification for a credential, the performance of all students must be evaluated.

Option #1: If a statewide program of assessment is adopted, institutions should be evaluated on the basis of student data which are summarized. Individual student progress should not be directly affected by the results of any assessment.

A performance evaluation program should only be used to evaluate and compare institutions. Individual students should not be directly affected by the results of examinations. Information which is generated should be used for discussion about the intent of the program: the outcomes produced versus the outcomes desired. Individual students should not be held back as a result of their assessment results.

Advantages

* Student results would be summarized to provide institutions and their publics with general information about the institution's programs.

* Since sampling procedures could be used, the cost of the assessment program would be reduced.

Disadvantages

* Individual students would not benefit directly from the results of an assessment program. Students would not have the advantage of remedying deficiencies identified by the assessment process.

* Since students would have no direct, personal stake in the outcome of the assessment, they may not participate to the best of their ability. The result could be an inaccurate measure of institutional performance.

Option #2: If a statewide program of assessment is adopted, individual students should not be permitted to enter the next educational level until they demonstrate the required proficiency.

Individual student evaluation should determine whether a student is prepared for the next educational level. Assessment results should assist students in acquiring the skills necessary for progression and for screening those who are not prepared for the rigor of each subsequent educational step.

Advantages

* All students are provided with feedback about their performance. Students who are in need of additional skill development are able to take the necessary courses to improve their preparation.

* Spaces in upper-division courses and graduate schools are not taken by students who are unable to gain the most from those experiences.

* Students would take the assessment seriously, producing a more accurate measure of institutional performance.

Disadvantages

* The assessment employed may not be the best measure of a student's ability to be successful. Some tests may deny access to students who do not test well.

* Since sampling procedures could not be employed, the cost of the assessment program would be greater.

WHO SHOULD BE RESPONSIBLE FOR PERFORMANCE EVALUATION?

A final question concerns who should design the evaluation. If institutions are permitted to design their own systems, it may not be possible to compare the performance of institutions. On the other hand, a statewide system applied to all institutions may not be sufficiently sensitive to differences among institutions.

Option #1: Each institution should be responsible for developing and implementing a comprehensive evaluation plan.

This option would permit each institution local autonomy in the design and implementation of an evaluation plan. Because each institution would have the flexibility of selecting what performance would be measured, local decisions would dictate what, if any, student assessments are incorporated in the plan and how they are used. The Higher Education Coordinating Board would be responsible for monitoring the institutions' plans.

Advantages

* Because the plan would be designed by institutions, it is likely to receive faculty support and the results utilized to improve the curriculum.

* A performance evaluation plan designed on a local basis is likely to be tailored to the institution's particular role and mission.

Disadvantage

* Performance evaluations which are locally based are not likely to yield data which can be used to

compare institutional performance. Only internal comparisons over time would be available.

Option #2: A statewide performance evaluation program should be established for higher education institutions which allows comparisons to be made among these institutions.

A statewide performance evaluation program would provide information about all higher education institutions. The Higher Education Coordinating Board would be responsible for designing a program which provided common information about how students perform at each of the institutions.

Advantage

* Common information will be available which will permit comparisons to be made among institutions.

Disadvantages

* The performance which is assessed may be so general that comparisons may be of very little value.

* Institutional personnel may be suspicious about how the results of a performance evaluation program are used. Because of this suspicion, a faculty and staff could undermine the effectiveness of a statewide program.

Option #3: A statewide performance evaluation program should be adopted. However, each institution would be able to add to the program to meet local needs.

The Higher Education Coordinating Board would be responsible for implementing a statewide performance evaluation

program. However, each institution would be able to add to the basic elements of the plan to evaluate unique program offerings or provide additional information which a particular institution would like to collect.

Advantage

* Comparative data would be available about the institutions, yet each institution would also have the flexibility of designing additional evaluation components to fit the individual needs of the institution.

Disadvantage

* Faculty may not be supportive of the overall program because of the lack of local autonomy in some areas of performance evaluation.

ATTACHMENT

THE EVALUATION OF INSTITUTIONS OF HIGHER EDUCATION IN THE STATE OF WASHINGTON

A REPORT PREPARED FOR THE HIGHER EDUCATION COORDINATING BOARD OF THE STATE OF WASHINGTON

FRED E. FIEDLER
University of Washington

GERALD M. GILLMORE
University of Washington

ROBERT M. THORNDIKE
Western Washington State University

DOUGLAS J. ALLMON
University of Washington

This report was prepared under contract with the Higher Education Coordinating Board, Olympia, Washington by a committee consisting of Fred E. Fiedler, Professor of Psychology and Adjunct Professor of Management and Organization (Chair), Gerald M. Gillmore, Director of the Educational Assessment Center, University of Washington; Robert M. Thorndike, Professor of Psychology, Western Washington State University, and Douglas J. Allmon, Research Associate at the University of Washington. Consultants to the project were Nathaniel Gage, Professor of Education, Stanford University; Wilbert J. McKeachie, Professor of Psychology and Associate Director, Center for Research in Education, University of Michigan, and Peter Evell, Director, National Center for Higher Education Management Systems, Boulder, Colorado. We are also indebted to Judith Fiedler, Manager, Center for Health Studies, Group Health Cooperative of Puget Sound, for her contributions to this report.

INDEX

Page

Executive Summary.....3

Introduction.....6

Testing.....8

Value-Added Measurements.....10

Assessing Alumni.....12

Performance in Graduate and Professional Schools.....13

Appendix A: Testing.....18

Appendix B: Value-Added Testing.....28

Appendix C: Assessment of Alumni.....35

Appendix D: Sample Questions.....42

Appendix E: Expense Considerations.....50

Appendix F: Reviever Comments.....54

Appendix G: Reference List.....62

THE EVALUATION OF INSTITUTIONS OF HIGHER EDUCATION

IN THE STATE OF WASHINGTON

EXECUTIVE SUMMARY

This report was commissioned to assist the Higher Education Coordinating Board of the State of Washington to provide technical assistance in implementing its decision to establish a system for assessing institutional performance of the state's four-year college programs. In particular, we address Question 4 of the state's Master Plan, dated September 1986:

> "How should the performance of higher education institutions be evaluated and by whom?"

The report is based on a review of the literature on the assessment of educational programs, and personal and telephone interviews with researchers and administrators involved in educational program evaluation. Our recommendations are here briefly summarized.

1. All sources agree that the assessment of performance of educational institutions is needed, feasible, and difficult. We fully concur with this judgment.

2. Academic institutions perform a multiplicity of complex functions for the student, society, and the state. Adequate assessment requires multiple measures. The institution's task to provide a general education for the student requires different measures from that of teaching cognitive skills, preparing the student for postgraduate training, or for taking his or her place as an economically productive citizen who contributes to the social and political life of the community.

3. Tests of the student's general knowledge and thinking skills can assess the adequacy of the institution's curriculum during the first two years of college. Tests of this nature, roughly tapping "what every educated person should know", are best administered at the end of the second year. They are available from commercial test development organizations, and can also be developed or modified to reflect the specific emphasis of the state's institutions of higher learning.

ACT Comp

4. In-depth tests of knowledge in the major fields of concentration are commercially available for many areas (e.g., Graduate Record Examination). Alternatively, core knowledge tests for graduates in a particular discipline can be developed by faculty from the different state schools. The latter is considerably more difficult and costly, considering the multiplicity of major fields of concentration offered in this

state.

5. The assumption that a student must gain from attendance at college is basic to the public's support of higher education. Measuring the amount of gain presents unexpectedly difficult methodological and statistical problems discussed in detail in Appendix B. The "Value-added" concept is attractive on its face, requiring that parallel tests be given at an early and a later stage in the student's college career. However, subtracting the first test score from the second leads to statistically flawed results, and we strongly recommend against its use. Achievement tests, using appropriate statistical analyses, serve the same purpose in a more defensible fashion.

6. The primary goal of college is to prepare the student for subsequent work and life in the community. The effects of college on the individual's subsequent career choices, contribution to the economic and social life of the community, and job and life satisfaction, can be assessed to an acceptable degree by random sample mail or telephone surveys. Earlier tests of these kinds have generally shown that college graduates are able to earn more, contribute more to the community's social and political life, and report higher job and life satisfaction. It is difficult to determine, however, to what extent the benefits of a college education are due to increased competence and knowledge, or simply to having a diploma. Development of a positive attitude toward sound thinking and learning may also be an area to examine among alumni.

7. Comparisons between schools and programs present major problems since the different schools have different missions and serve different self-selected segments of the population. While "peer group" schools or programs can be compared to some extent, measurement against an acceptable set of standards provides a more satisfactory method in most cases. Some comparisons among institutions are possible by using appropriate statistical methods for controlling differences in e.g., admissions standards. Unfortunately, these statistical adjustments may be difficult to interpret to the public.

8. Administrators and researchers agree that any method of academic evaluation, must have the full trust and cooperation of the faculty. We recommend that the evaluation involve the faculty and administration of each college throughout the process. The faculties of the various schools be charged with choosing the topics and specific test items in all program evaluations order to assure that these reflect the current status of the field and the orientation of the various faculties.

9. We strongly urge that the Board proceed with caution in implementing any evaluation system. Unexpected problems are bound to arise and may jeopardize the success and acceptance of the entire enterprise. A pilot test over a one- or two- year period essential in bringing to light possible weaknesses in the program's concept and the logistics of its implementation. It is

important not to trivialize the testing endeavor by measuring abilities, skills, and knowledge that ought to have been acquired in high school.

10. We recommend to the Board begin a program of evaluation on a relatively simple basis, consisting of (a) a general knowledge test administered at the end of the second year of college, and (b) alumni questionnaires administered to an appropriate sample of graduates two years after graduation. A model evaluation program might be submitted to the several colleges and universities as a working basis for further development.

INTRODUCTION

A. Structure of this Report:

The first part of this report contains a relatively short and nontechnical discussion of the major questions to be considered in the development of a program for evaluating the performance of the four-year programs of state universities and colleges. Specific sections deal with the adequacy and usefulness of available testing programs, the Value-added problem, and measures based on alumni and employer reactions.

In-depth discussions of educational achievement testing, value-added approaches, and post-graduation measures are presented in Appendices A and B. Appendix C presents the method and logistics, considerations applicable to post-baccalaureate measurements. This section also discusses cautions and possible pitfalls brought to light in the literature and discussions with our informants. These need to be considered in order to make the evaluation programs successful. The final section consists of discussions and evaluations of the proposed methods by experts in the field who served as outside consultants.

B. Who is the Consumer

The measurement of institutional performance depends on identifying the intended consumers of the State's higher education system.

> If the primary consumer is the student, is the educational experience intended to (a) lead to a richer, more satisfying and fulfilling life; (b) open opportunities for further professional training; or (c) prepare for greater employment opportunity?
>
> If the consumer is the potential employer, should colleges teach intellectual skills and technical knowledge to make the graduate more employable and productive?
>
> If the consumer is the State or the public at large, will the college education contribute to (a) the productivity and welfare of its citizens and (b) their ability to participate in the social and political life of the community?

Depending upon the answer to these questions, we shall need different products and utilize different measures. Table 1 shows the various potential consumers and the corresponding products of the educational system:

Table 1

Consumers and Potential Products of Higher Education

Consumer	Product	Sample	Criteria
Student	Knowledge Intell skills	College students	Achievement tests "Value added" competancy tests
Student	Employability	Recent grads 2 years later	Job Satisfaction
Student	Enrichment of life	Prevous grads of 4-yr instit. +2 - 5 years	Life satisfaction Job satisfaction Integration in community
Grad School Profl schools	Eligibility grad trng	Graduates 1 - 4 years	Students admitted and performance in grad school
Employer	Job relevant skills	Business orgns +2 - 5 years	Economic benefits, Job skills
Public State	Contribution to public good	Recent grads +2 - 5 years	Participation in public affairs Contribution to community

D. Intended and Unintended Consequences of Evaluation

It is inevitable that the measurements adopted by the HEC Board will have major consequences for the governance of the four-year institutions of this state. Some of these will be intended, others will be unintended or undesirable. It is important, therefore, to forecast, insofar as possible, the probable positive and negative consequences of each measurement approach. We shall attempt to do so, although all potential consequences of an evaluation system can rarely be foreseen.

Ideally, a criterion should assist an organzation to achieve desired outcomes. If the organization has, for some reason, veered off course, the criterion measure should ring appropriate warning bells, or better yet, provide a servo-mechanism that will automatically make the needed corrections. One example of this type might be a rule that certain elective courses can only be taught if at least 20 students registered. As a result, poorly tought courses, or courses of no interest to students

automatically are eliminated.

E. Diversity and Comparability of Institutions

In establishing a variety of institutions and educational programs, the State obviously intended to offer students a variety of educational experiences. Criteria must, therefore, allow for the different philosophies that guide the State's diverse educational programs. The most equitable method for evaluating schools appears to be a comparison with that school's peer group, i.e., schools of comparable size, student population and resources. The HEC Board might consider a systematic effort to identify schools in other states that are part of an institutional evaluation program. Where comparisons are to be made among the schools of the state, appropriate statistical controls must be used to assure the fairness of the comparisons.

F. Consequence

The most frequently utilized method of evaluating performance in higher education is to assess student knowledge at various points in their college careers. Strategies of assessment vary, intentionally or unintentionally, in the extent to which they "invade" the domain and prerogatives of the faculties and administrations of the various state institutions. In these terms, a highly invasive approach prescribes admission standards, courses, and examinations; it would mandate conformity with HEC Board or legislative standards. The least invasive approach sets broad goals but leaves to the institutions the choice of methods and maximum freedom to express their uniqueness.

If the Board mandates certain examinations (for example, in history, trigonometry, or botany) it will influence institutions to prepare students to meet these demands. If the Board evaluates the jobs students get after graduation, then institutions are likely to direct additional resources to vocational guidance and placement centers. Finally, if the Board mandates assessment at any point in students' education, institutions will respond accordingly, which in turn will affect the mix of required courses. The following section presents short and relatively non-technical discussions of testing, value-added approaches, and alumni surveys.

TESTING

Testing activity for assessing university programs and for assessing student growth and competence can be classified into two types. General academic skills testing deals with a broad spectrum of abilities and understandings and is oriented around what all students should achieve. We use the term to include components of "liberal education" such as art and music appreciation, as well as the ability to think logically, write coherently, and do simple mathematics. In-depth testing, in contrast, is oriented toward testing students within their specific majors. It is based on the widely held notion that a

graduating student should hold in-depth knowledge in at least one discipline and that an important purpose of a university is to foster that knowledge.

General academic skills can be tested upon entry, and/or at some point during the college career, often mid-way. A major purpose for entry testing is the placement of students into appropriate classes. A related purpose is to provide information for curricular planning, specifically for structuring introductory-level courses. (Entry testing can also produce base-line or "before" data, which then can be compared to later test results to assess gain.) Testing at some middle point in the college career can be done for evaluating students and/or programs. These test results can provide data, either in and of themselves or in comparison to entry data, to evaluate and revise the curricular offerings and modes of teaching. The so-called "rising Junior" concept demands that students achieve some minimal score on such an exam before being allowed to do further non-remedial work toward graduation.

In-depth testing is typically done at or near the end of the senior year. Its results can be used to evaluate students, requiring a passing score for graduation, or to evaluate programs. For the latter evaluation, the assumption is usually made that students begin college with essentially no in-depth knowledge of their major discipline, and hence pre-testing is not needed.

In evaluating different tests or testing orientations, we suggest the following seven issues should be addressed:

A. Test Availability. Are tests available which will measure the particular aspect of the institution's mission and/or can tests be developed locally?
B. Cost. What are the costs associated with the various options, both out-of-pocket and in terms of human resources?
C. Reliability. Will the given test provide consistent measures?
D. Validity. To what extent are resulting test scores indicative of that which one wishes to measure as opposed, for example, to being an artifact of the testing method or based on irrelevant content?
E. Usefulness. To what extent do the test results yield information which can be used to improve instructional programs and products.
F. Standards. Against what standards or criteria will the test results be compared?
G. Sampling. Can testing be applied to random or stratified samples of students or must all students be tested?

Table 2 contains a brief summary of how testing for each of the six purposes mentioned earlier and listed in the top three rows of the table relates to each of these issues. To further summarize, these purposes can be ordered in terms of the anticipated difficulty in achieving successful implementation. The easiest program seems to be

TESTING SUMMARY

Table 2

TEST TYPE	BASIC ACADEMIC SKILLS				IN-DEPTH	
TIMING	Beginning of Freshmen Year		Prior To Junior Year		At the End of the Senior Year	
PURPOSE	Placement	Evaluation of Introductory Courses	Evaluation of General Curriculum	Advancement to Junior Status	Eval of Departments Offering Majors	Requirement for Graduation
AVAILABILITY	Readily available for many subjects		ACT COMP best commercially available test battery. Local development very difficult.		Few acceptable national exams are available. Local development not difficult.	
COST	Costs may be passed on to students. Some costs already borne by colleges.	Some expense: cannot pass costs to students unless part of placement program.	ACT COMP can be quite expensive, depending upon how much of battery is used. Local development is expensive in terms of necessary involvement of large number of faculty.		National exams tend to be expensive. Local development relatively less expensive because exams are subject matter specific. Some expense for for outside consultants and for manual grading of "essay" items.	
RELIABILITY	Not a serious issue for standardized tests. Locally developed tests will typically have good reliability.		Some problem since items are chosen from large number and variety of potential items.	Major concern given decision made as result of test. Will 2 versions give same result?	Needs to be checked carefully in pilot tests of locally developed exams, especially if graduation is dependent upon passing the exam.	
VALIDITY	Do pilot testing to see if test correlates with course grades.	Faculty assessment that test item content is appropriate.	Faculty assessment that test item content is appropriate. Likely to be difficult to satisfy all parties.		Mainly an assessment of the item content and required skills. Suggest addition of outside faculty and community members to broaden viewpoint.	
USEFULNESS	Can help maximize student academic experience.	Useful for revising particular courses.	Difficult to use test results to make specific changes.	Only useful if remediation and re-testing is built-in.	Useful for evaluating and for guiding curricular decisions.	Only useful if remediation and re-testing is built-in.
STANDARDS	Compare to later course performance.	Assessment of expected performance given curriculum.	ACT COMP would provide norms. No good standard for locally developed	Almost impossible to develop justifiable standards for advancement. test.	No very good standard against which to compare performance.	Very difficult to determine justifiable minimum score for graduation.
SAMPLING	All students who plan to take appropriate courses.	Can test a sample of students.	Can test a sample of students.	All students.	All students in major. Can sample majors on rotating basis.	All students

placement testing upon entry, which is already accepted and used on most of the state's campuses. Furthermore, many tests for this purpose are in existence, and validation can be accomplished in a straightforward manner by relating test scores to later course performance and by faculty assessment of item content. Related evaluations of introductory course content with an eye toward possible curricular revision is a direct spinoff of placement testing.

Of middle difficulty is the in-depth testing of seniors in major areas of concentration. Few off-the-shelf tests are available for this purpose, and those that are available tend to be expensive. However, local development of tests by faculty within these departments, aided perhaps by faculty from other campuses and appropriate community members, is clearly feasible. In fact, the test development (or selection) process in and of itself would be valuable in causing faculty to address their department's goals, and test results ought to provide valuable data for addressing curricular and teaching issues. However, adding the requirement that seniors pass the test in order to graduate should be approached with great caution, given the difficulty in establishing fair and reasonable standards, and given the importance of such a decision on human lives. Variability in individual test anxiety levels also will affect examination scores and perhaps the future careers of individuals.

The testing of basic academic skills sometime midway or later in the college careers of students presents some problems. The valued diversity of the general education students receive requires faculty agreement on what should be included. Furthermore, many of the general skills which we find most important are the hardest to measure validly. These include, for example, generalizable problem solving ability. The latter also seems least likely to yield useful results, because it lacks a clear connection to definite instructional improvement activities. Even more important, it seems nearly impossible to use resulting individual test scores to judge whether an adequate general education has been attained such that a student can advance to specialized study. Again, the importance of such a decision requires very careful preliminary study.

Finally, continuing fine-tuning of measures should be encouraged. Both objective and subjective (open-endend) formats may be used, depending on the degree of validity required.

VALUE-ADDED MEASUREMENTS

A. Introduction.

The term "value added" is a currently popular one to describe attempts to determine what changes have been accomplished by a college education. The phrase implies the assessment of changes in students, changes that are assumed to be desirable. It derives from a type of tax imposed by many European countries to reflect the change in value of goods at various stages of the manufacturing process. A college education is thus seen as

analogous to a step in the process of manufacturing citizens, and the society that supports the college may use the valued-added measurement to determine whether it is getting its money's worth.

There are two basic approaches to value-added assessments. One is to specify a desired change or class of changes and determine whether these have been accomplished. To a certain extent, the determination of what changes are desirable and the responsibility for determining whether the desired changes have taken place has been the responsibility of the faculties of the state's universities and colleges. This is accomplished in the setting of the curriculum and the approval of courses that make up the curriculum. Given the diversity of the State's institutions of higher learning, the present state of knowledge about the effects of a college education, and the great diversity of paths open to students as they pursue their educations within a given institution, this alternative is not feasible to assess value-added state-wide.

The other alternative is to specify broadly some areas of interest, such as cognitive or attitudinal changes, and, measuring at two or more points in time, to determine what changes actually take place, reserving judgment about these changes until they have taken place. Any assessment of changes wrought by Washington public higher education will have to be of this kind until much more is known about the processes and consequences of higher education.

B. Possible Topics and Methods

1. Topics

The first step is to select the topic to be assessed and what general types of changes are desired. There are three broad areas that present possibilities; changes in cognitive skills, changes in social skills, and changes in attitudes and beliefs. All three are reasonable areas for study, but it is chiefly in the first two areas that it is reasonable to speak of improvement without imposing a subjective value system. That is, most people would agree that cognitive growth, the ability to solve progressively more difficult problems or possessing a larger fund of knowledge, is a desirable objective of a higher education. Likewise, most people would agree that becoming more socially competent, more able to deal in an effective way with other human beings within our culture, is a desirable change.

However, in the area of attitudes and beliefs, definition of desirable outcome becomes more subjective. Is it desirable that students become more liberal as a result of going to college, or should they become more conservative? Should they become secular humanists or should their belief in divine powers be strengthened? Reasonable people can clearly disagree on answers to these questions and hence these are not appropriate subjects to address in terms of "value added." However, they are changes that are important to the person and are effects of which

the colleges and universities should probably be aware. These changes in attitudes toward learning and studying should be measured.

The determination of change requires that information about current level of functioning be available at at least two points in time, at the beginning of the "treatment" and its end. It may also be desirable to monitor progress by taking additional measurements at various points in the process.

It is important to pay careful attention to the role of value at this stage of the investigation. Policy makers justifiably want evidence about how well resources are being used, but too little is known about the consequences and correlates of the college experience to permit making inter-institutional or inter-program comparisons. Until a much wider base of information has been accumulated, it is probably unwise to use gain scores for comparative purposes. Studies should describe what exists. Later it may be possible to address the question of what should be.

2. Measurement Methods

There are two basic measurement methods that can reasonably be used to assess gains in a large number of students in a wide variety of curricula. These methods are tests and self-report questionnaires. The former are appropriate for the cognitive areas, while the latter are appropriate for the topics of beliefs and attitudes. The measurement of social skills, if attempted, is also likely to be undertaken with self-report methods because standardized measuring devices are not now available in this area and the problems of developing such measures are very substantial. See Appendix 2 for more detail.

ASSESSING ALUMNI

In the final analysis, the benefits of higher education must be judged by the benefits the student derives after graduating from college: How has college affected the individual's life?

A number of investigators have examined the effects of college education on alumni lifestyle, performance as productive citizens, and contributions as useful members of society. Included in these studies are measures of the extent to which alumni have achieved employment appropriate to the fields for which they have been educated, earn incomes relatively higher or lower than selected criterion levels, use the community's cultural resources, and assume civic responsibility. The latter includes voting, involvement in the political process, and participation in civic enhancement endeavors.

Alumni questionnaires also measure the extent to which alumni express satisfaction with their lives, careers, and educations. Ideally, a university graduate and citizen would

demonstrate some sophistication in decision-making which affects others. Several questionnaires are available for measurement of moral development among alumni. While not sufficiently reliable for evaluating particular individuals, they may be useful aggregate or average measures for establishing whether college has had the desired effect on students.

Comparison of survey results from alumni of schools presents considerable problems because of the influence of differences in motivation, socio-economic background, economy, and demography by region. These influences will affect job satisfaction, earning capacity, and access to cultural resources of the state. We can, however, evaluate the degree to which individuals live productive and successful lives according to prevailing community standards.

To make an adaquate evaluation of lifestyle and productivity among alumni will require simultaneous use of attitude, opinion, and survey techniques. A census of the entire population would entail considerable expense. However, use of carefully designed and administered sampling provides accurate and generalizable findings at a relatively lov cost.

The two most reasonable time periods after graduation appear to be two years and five years. After two years, the individual's life is still directly affected by experiences in college. This is most clearly the case for students who entered professional and graduate programs. However, even for most alumni, the knowledge and the intellectual and social skills acquired in college have a considerable influence on the individual's economic and social functioning level.

The methodology of survey research is well known, and the surveys can be conducted under contract by a private organization, under direct supervison of the HEC Board, or by each of the uiniversities and colleges. The actual size of the sample does not have to be very large unless specific subsamples of students (e.g., by department or engineering college) are to be evaluated. To ensure adequate representation of each academic discipline it is possible to sample graduatates of specific academic departments in successive years. (See Appendix C).

PERFORMANCE IN GRADUATE AND PROFESSIONAL SCHOOL

One important function of undergraduate education is to prepare the student for professional and graduate training. The evaluation of the four-year program is less difficult in this respect than in the others discussed above. The basic questions are, How good is the school which admitted the student? How well does the student perform in comparison to those from other four-year college programs?

In this case, comparison with peer institutions is relatively easy. Two of several measures seem appropriate. The

first evaluates the quality of the school to which the student is admitted, using such indices as the ACE rating of the department, the average GRE scores and GPA of applicants, etc. The second index is a rating of the student by faculty of the graduate program. A third index can be obtained later on by rating the desirability of the student's first job after graduation, and five years after graduation. It should be noted, however, that these indices are strongly influenced by self-selection of the student to an undergraduate program and by the selectivity of the school to which the highschool student applies.

Table 3 lists various measures of potential use in the assessment of educational institutions.

Table 3

List of Potential Educational Criterion Measures

1. Input (Not considered)
   - How good are students who apply? How attractive is the school?
   - How good are the students the schools admit ACE or SAT scores upon entry, GPA from high schools?
   - How good are the faculty members attracted to the school Schools they come from, competing offers, student ratings?

2. Process variables (not considered)
   - How good are the courses according to student ratings
   - How many students register in various classes
   - How many students graduate
   - Student grades in various colleges and programs

3. Output measures-
   Measures of student achievement after graduation
   a. - How desirable are the jobs students accept upon graduation
   b. - How good are the graduate schools that accept them
   c. - How well do students perform in graduate schools

4. How much do students know at graduation?
   a. - General achievement tests (ETS, Psych Corp, etc.
   b. - Achievement for each major - tests made up statewide, corrected for socio-economic status and geographical area, etc.

5. Value Added Criteria: How much have students learned since entering college?
   - Gain scores: differences in scores on achievement tests (cognitive domain) between entry and graduatio

6. Reputational criteria
   a. - Ratings by other schools
   b. - ACE ratings
   c. - Ratings by graduate schools

7. Has college contributed to the "good life"?
   a. - How satisfied with life are graduates of various schools?
   b. - To what extent do graduates participate in the cultural life of the community? Theaters, movies, books, lectures, musical events, TV programs, sports events? Do they participate in civic affairs and the political process?
   c. - How satisfied are graduates with their jobs, their salaries, their social and family life?

Potential Pitfalls and Expected Criticisms

It is virtually inevitable that any program of evaluation chosen by the Board will be subjected to severe criticism and attack. For this reason, we discuss a number of points most likely to be raised in critical discussions.

Most indices of accomplishment really cannot be interpreted without some comparison groups, but the identification of an appropriate comparison group represents a major problem in assessing college programs. We can determine without difficulty whether students have reached a given level of reading skill, acquired a certain amount of knowledge, and/or express satisfaction with life or with their work. But, would they have done equally well without a college education? With whom can we compare these individuals?

If we use as our comparison group those who have not attended college, then we need to consider the major differences that exist in (1) educational and socio-economic background of the families, (2) motivation and ability, (3) and opportunities. If we compare students from state and from private institutions, we again have the problem of selection. Those in private institutions are likely to come from families differing in income, professional, and religious backgrounds. These differences imply a more conservative and more religiously-oriented population sample. In other words, there is a major self-selection factor operating. If we use those who have entered but not completed college, we again have a problem in accounting for differences in self-selection, motivation, and abilities.

As already pointed out, comparison is least open to attack when the comparison group consists of schools in the same league. Use of other comparisons would raise problems which must be dealt with by using statistical techniques that are relatively difficult for the public to understand.

If we take gain scores, we have different problems. Those who enter with a high score typically do not gain as much as those who enter with a low score. Those who enter late will progress at a different pace than those who are early entrants. Again, suppose that students from a school gain 20 points on the average. What does this finding mean? Is it large or small? Various statistical methods (e.g., analysis of covariance) may be able to cope with the problem of gain in a manner that can be defended as methodologically sound. At present, no universally acceptable method is available for the purpose of indicating gains in educational programs, and approximations will have to serve for the time being.

Surveys of alumni indicate the degree to which (a) the college graduate contributes to the social and economic life and well-being of the community and (b) the alumni are satisfied with what their college education provided. The disadvantages of

survey research are the cost and the difficulty of identifying exactly which benefits derived from college education. Furthermore, the question how a college education should affect to a person's earning level, occupational success, use of libraries, attendance at theaters, or appreciation of music is bound to generate controversy. Whether effects of education weaken over time is a question for further research.

It is inevitable that there will be a tendency to compare the state's institutions on various measures. These comparisons, no matter how well intended, will have several undesirable results. For example, it may induce colleges to give preference to high school students who already have a good deal of relevant knowledge, thus discriminating against students from disadvantaged minority groups.

The faculty and/or administration of an institution may also feel that it is not the school's mission to provide a general academic background of the type evaluated by commonly-used tests. Faculties at engineering or agricultural schools might not wish to concentrate on general education. Yet evaluation based on criteria of general education may result in adverse evaluation of their respective schools.

The argument against use of prescribed measures is likely to be that an assessment of this nature will homogenize the colleges, force an undesirable degree of uniformity, freeze the types of courses that will be taught or the educational experiences students will obtain in the various schools. This problem can be overcome by permitting schools to develop their own criteria for evaluating their performance within reasonable limits. For example, the faculties from each of the schools might be invited to select tests or a pool of 100 to 500 test items appropriate for its student body, and the examining body could then select at random a set of items for any one test. This plan would let each school set its own standards while these standards would be subject to evaluation by the HEC Board, its consultants, or the faculties of other state schools. Setting very low or inappropriate standards would tell its own tale.

APPENDIX A

TESTING

(Gerald M. Gillmore)

I. Basic Principles

There are several basic principles which run throughout the literature on the assessment of college programs. These principles are relevant to all aspects of testing in service of assessment and therefore will be briefly discussed at the outset of this section.

A. The ultimate purpose of assessment is improvement of the quality of education. In planning and evaluating assessment programs, one's foremost consideration should be whether the programs foster improvement in the education received by the students who are enrolled at the institutions. The assessments by themselves are unlikely to have this result, but if done properly can facilitate such a result through the process of thought which accompanies the planning of the assessments (e.g., values clarification) and through changes made based on the results of the assessments (e.g., curricular revisions).

B. Each institution must determine its own unique mission and build assessment programs appropriate to that mission. Assessment programs are expensive in terms of human resources and actual cash outlays. We cannot afford expenditures on assessment processes which are not germane. Furthermore, information which is not oriented to the institution's mission will not be used to effect meaningful change.

C. The faculty from each institution must be involved in the planning, implementation, and interpretation of the assessment programs and resulting data. This principle flows directly from those above. If programs come from the state level or from the institution's administration without involving faculty, then the results will not be used to effect meaningful change. Rather, they will be viewed as intrusive and irrelevant. Moreover, the mission of the institution is likely to be misspecified without serious faculty input.

II. Test Types

There are three types of tests which are associated with the assessment of the programs of higher education: minimum competency, general academic skills, and in-depth understanding.

A. Minimum Competency. This type of testing usually deals with certain skills or abilities which are judged to be essential to successful performance in everyday life. Examples of such abilities include balancing a check book and looking up a telephone number in a directory. We judge this type of test to be irrelevant to the subject at hand for several reasons. First, it is commonly agreed that such outcomes are the proper province of the nation's secondary schools. The subset of the population continuing their education beyond high school should be expected to have mastered minimal competencies, so defined, upon entry. Secondly, minimum competency often translates to "adequate" competency and no institution of higher education wishes to so reduce its mission. Finally and relatedly, considerations of only minimum competency tend to trivialize higher education such that faculty feel uncomfortable in their support and cooperation in the

assessment thereof.

B. General academic skills. General academic skills testing can be differentiated from minimum competency testing in that it deals with a much broader spectrum of abilities and understandings, and it is less practically oriented. We use the term "general academic skills" very broadly to include components of "liberal arts" education such as art and music appreciation as well as the ability to think logically, write coherently, and do simple mathematics.

Relative to minimum competency testing, general academic skills testing tends to be less interested in the achievement of some minimal cutoff score by individual students, Nonetheless, institutions can choose to set minimum cutoff scores for entry or advancement. Of course, problems with determining appropriate minimal levels must be recognized and will be discussed subsequently.

General academic skills as we define them can be tested prior to entry, upon entry, and/or at some point during the college career. Testing prior to entry is done for admissions purposes using a standardized test such as the Washington Pre-College Test or the Scholastic Aptitude Test. These tests try to assess aptitude for college level education rather than achievement, though in practice the distinction is blurred. We view this type of testing as well-worked out and not directly relevant to assessing college programs. Testing upon entry may be done for several reasons. One reason, important to assessment of the impact of the college's instructional program, is to produce base-line or "before" data, which then can be compared to later test results to assess gain. Another reason for entry testing is the placement of students into appropriate classes. A third reason is to provide information for curricular planning.

Finally, general academic skills testing at some middle point in the college career can be done for evaluating students and/or programs. From a program assessment perspective, such test results provide data, either in and of themselves or in comparison to entry data, to evaluate and revise the curricular offerings and perhaps the mode of teaching. In terms of evaluating the progress of individual students, the so-called "rising Junior" concept demands that students achieve some minimal score on such an exam before being allowed to do further non-remedial work toward graduation.

C. In-depth Testing. This type of testing can be contrasted with general academic skills testing in that the former tends to test broadly and uniformly across all students while the later is in-depth and oriented toward students within specific majors. It is based on the widely held notion that a graduating student should hold in-depth knowledge in at least one discipline. In-depth testing can be used to evaluate students, requiring a passing score for graduation, or programs. For the latter, the assumption is usually made that students begin college with essentially no in-depth knowledge of their major discipline, and hence pre-testing is not needed.

III. Issues

In evaluating different tests or testing orientations, there are many issues, some psychometric, some practical which must be addressed. We briefly enumerate these issues as follows:

A. Test Availability. Are tests available which will measure the particular aspect of the institution's mission or will tests have to be developed locally? Clearly, standardized tests produced by reputable national test companies will tend to have psychometric properties superior to locally developed tests; however, the former will also be less sensitive to particular program idiosyncrasies. This fact is exacerbated for in-depth testing where each major requires its own examination.

B. Cost. Some proprietary tests are prohibitively expensive to administer to large groups of students. Even less costly tests can be quite expensive when their cost is multiplied by many students. Locally produced tests, on the other hand, can represent less "out-of-pocket" expenditures but can take their toll in use of valuable human resources.

C. Reliability. Reliability may be considered the minimum competency principle of testing; that is, the test must measure consistently. Two more subtle issues relate to reliability within the particular application of program evaluation. First, tests may be highly reliable because they measure a very narrow content domain. Such reliability sacrifices more important outcomes. Secondly, tests which yield reliable scores for comparing individuals do not necessarily yield reliable scores for comparing groups.

D. Validity. This complex concept is critical to the entire enterprise, referring roughly to the extent to which the test scores are indicative of that which one wishes to really be measuring. Validity is the single most important quality a test can have. Without validity, decisions based on test results will be incorrect whether they deal with programs or with people. Typically, three types of validity are identified. Construct validity is based on the underlying theoretical structure associated with the measure. (In a sense all validity is basically construct validity because what we measure is based on our theoretical notions of what we should be measuring.) Predictive validity refers to the extent to which a test correlates to a measurable criterion, such as course grades. Content validity refers to the extent to which the content of the measuring device matches the content of that which one wishes to measure. For example, a multiple choice test has questionable content validity for measuring writing ability. Tests of simple addition have little content validity for measuring mathematical competence.

E. Usefulness. It is not enough that a test be both valid and reliable. It should also be useful. For example, valid and reliable tests of general exist, but such tests would not be useful for evaluating college programs, because presumably intellence is an innate property. In this context of college assessment, to be useful the test should yielding results which can be used to improve instructional programs and products.

F. Standards. Against what standards or criteria will the test results be compared? Often there will be no natural comparison. For example, each major needs its own test for in-depth testing; thus, test results will be non-comparable across majors, unless national norms are available. But even with national norms, comparisons are suspect. Comparisons within the same major from one year to the next may be quite informative, however. Importantly, one has to always

balance the information gained by the given comparison with the validity of the comparison and the threat of the comparison.

G. Sampling. If testing is being used strictly to evaluate programs, often random or stratified samples will suffice. The results of a well-drawn sample will generalize accurately to the entire population with considerable savings in money and time. Testing from which decisions are made on individual students require that all students so affected be tested. Similarly, tests to evaluate smaller programs, such as specific majors, will also require examining all appropriate students.

IV. General Academic Skills Testing

The category of general academic skills tests will be discussed in the context of each of the seven issues enumerated above. Naturally this discussion will be far from exhaustive. Many of these issues will need to be pursued in much greater depth depending upon how the state higher education community decides to proceed. In particular the reader should recall that we posited several possible purposes for general academic skills testing. Relating to decisions about individual students, this type of testing can provide data for initial course placement and can serve as a gate for advancement to Junior status. In regard to program evaluation, general academic skills testing can provide information for curricular decisions regarding introductory level courses, can provide a baseline for later comparisons, and can measure students' competence after two or more years of college work. Clearly, a single test or unified battery of tests would be unlikely to serve all of these purposes.

A. Test Availability. Tests for initial course placement and introductory-level curricular decisions are readily available for many subjects. Over 4000 incoming Freshmen already take the Math Placement Test of the statewide Academic Placement Testing Program, which is a cooperative effort of the state's five Universities. Most community colleges have an ambitious course placement testing program. The Washington Pre-College test is used to make placement decisions at some state universities. On the national scene, the New Jersey College Basic Skills Placement Test Program is exemplary in testing and placing New Jersey students, particularly in remedial courses where needed. Tests of standard written English and reading skills , including writing components, are readily available from the Educational Testing Service (ETS) and other testing companies. Foreign language placement testing is done with locally developed tests at UW -- no acceptable national exam seems to be available for foreign languages.

Tests to serve the remaining purposes probably need to be more broadly focussed and more advanced. The most popular test battery for these purposes is American College Testing's College Outcomes Measures Program (COMP) battery. This battery measures three "process" areas: communicating, solving problems, and clarifying values, and three "content" areas: functioning within social institutions, using science and technology, and using the arts. The COMP is offered in three forms. The Composite Exam takes four hours to complete and includes questions requiring brief written responses, the writing of letters and memos, and brief oral presentations, along with multiple choice questions. According to ACT, this version requires a team of 5 faculty

a combined total of one hour to score the exam of each student. The Objective Exam consists of all multiple choice questions, takes two hours to complete, and measures the same six dimensions mentioned above. Finally, there is an Activity Inventory which assesses student (and/or alumni) involvement in various out-of-class activities relating to the six dimensions.

Other available tests tend to have been developed for other purposes but might be appropriate for measuring general academic skills. For example, the ETS College-Level Examination Program (CLEP) was developed as a means to give students course credit for knowledge gained from other than the classroom, but tests are offered in general areas such as English Composition, Humanities, Social Science and History.

Some institutions might prefer to develop their own tests. A major consideration in developing or choosing exams in this area is what content should be covered, how much of the test should be devoted to each area, and at what level of difficulty should it be covered. Developing ones own test forces these decisions to be explicit. However, assuming great breadth in these tests, test development would be a particularly difficult project involving a large number of faculty. It would ease the job if large item banks were available from which faculty could choose items rather than having to construct them. While we do not know of such, a large testing company like ETS might be a source, given the large number of items from previous tests which they have at their disposal. It should be noted that the test developed or choosen will have implications for restructuring the school's general eduaction requirements.

B. Cost. It is difficult to make definitive statements about costs for these exams without knowing the quantity and conditions of their use. Generally placement tests need not represent a major additional cost because many institutions are already engaged in placement testing, and the precedent of passing their costs along to students is set. On the other hand, if these tests are not used for placement but for curricular evaluation, passing costs along to students may not be justifiable.

Testing for more generally assessing academic ability would appear to be more expensive. The COMP Comprehensive exam, for example, costs $10.00 to $20.00 per examinee, depending upon how much of the package is purchased, if more than 1000 students are tested annually. In addition, 1 faculty hour is required for grading each exam. The Objective test costs only $6.00 per examine, again if more than 1000 are tested, and no faculty time is required for scoring. The Activity Inventory costs $2.50 per person. The CLEP tests cost $30.00 per examinee on an individual basis. Perhaps ETS would offer institutional rates if approached.

A locally produced test would necessarily be a highly labor intensive operation. A critical consideration in deciding to take this tack would be the value of the activity in and of itself in terms of forcing an institution to examine its mission. The result would likely be a more credible exam, but one which would be considerably more expensive unless the development costs could be amortized over many years of use of the test. Of course, there would have to be a mechanism to evaluate and up-date such a test, which implies some on-going expenditures.

C. Reliability. The reliability of tests for placement would not be expected to be problematic. Any well-developed test of fairly specific aptitudes or achievements will probably yield consistent results. The exception to this may be the direct testing of writing ability. Methods have improved to the point where scoring can be done with reasonable reliability. Less clear is the dependence of the writing score on the topic about which students are asked to write.

The reliability of gain scores, that is the difference between pre-test scores and post-test scores, is less definitive and will be covered in a later section on the value-added concept.

Tests of broad, general academic achievement offer an additional reliability-related concern. Such an exam by its nature contains few items relative to the vast number of items which could be asked. While students would be expected to be measured consistently with the same exam, careful consideration needs to be given to the question of whether results would change significantly if the test were comprised of different items, drawn from the same potentially vast set of items. This concern is particularly critical if advancement decisions are made concerning individuals.

D. Validity. Both predictive validity and content validity are important considerations in evaluating placement tests. Predictive validity is indexed by correlations between the test scores and later course performance. Most carefully written or chosen placement tests will satisfy this standard. Equally important is content validity which refers to the extent to which the test content reflects legitimate prerequisite knowledge needed for successful course performance. Content validity must be assessed by subject matter experts; it is not a statistical concern. Good placement exams should satisfy both types of criteria if they are to be used for placement and if they are to be used for evaluating the level of introductory course offerings.

The validity of tests of general academic skills which are not focussed on specific subjects is an incredibly complex, but absolutely vital consideration. To provide perspective on the question one needs to ask what is the meaning of a given score for a student or of an average score for a group of students. The complexity comes from two sources. First, the most important outcomes of a college education are those which are the most difficult to measure, e.g., problem solving skills, ability to write clearly, the ability to contribute to society, critical appreciation of the arts, etc. Easiest to measure are facts; but as important as they may be, no one would be satisfied with judging a student or an education program on the basis of the number of facts which are "known."

Secondly, the hallmark of American higher education is its diversity, both within and across institutions. General academic skills assessment would be much more easily measured if all students followed the same curriculum and were expected to attain the same results. Becuse this is not the case, and because we would not wish it to be the case, a test will have a variety of items to cover the breadth of general offerings. Even so, students following certain curricula might have an advantage. In general, pre-engineering majors will be expect to do much better on quantitative-type items than, say, those interested in the humanities and vice-versa for humanities-type

items. Does one hold both groups to the same standard?

It is our view that these issues are not yet fully resolved and that there is danger in "losing the baby with the bath water" unless the state higher education community proceeds very carefully. It seems crucially important not to trivialize higher education goals and outcomes, such as often has accompanied the behavioral-objectives movement in the public schools. It seems equally important not to jeopardize the healthy diversity found in higher education. In order for general academic skills testing to have validity, they must be developed over time with many intermediate data collection and evaluation steps before real decisions are made based on the data. This development must fully involve large segments of the professionals of higher education, the faculty.

E. Usefulness. Placement tests can be very useful in maximizing the students' academic experience through proper initial placement. They can also improve instruction in that they can offer assurance to teachers that all students have certain prerequisite knowledge. These tests seem also to be useful in evaluating curricula to fit the real abilities of students as they enter college, perhaps revealing the need to offer lower level classes, for example.

The decisions to be made on the basis of broad general tests are less clear. They may be able to highlight deficiencies in particular students, such as poor writing skills which can be remediated prior to graduation. They may similarly highlight weakness in the overall curriculum, e.g. is too little instruction or practice in writing. The ultimate usefulness will depend upon the extent to which the tests measure important outcomes which can be affected by curricular revisions in the case of groups of students and by remediation in the case of individual students.

F. Standards. Placement testing effectiveness rests on internal relationships between test scores and subsequent course success. Even evaluating the introductory level course offerings depends mostly on comparing expected and actual entering abilities and knowledge of students, and poses no severe problems in identifying appropriate standards.

Requiring a passing score on tests of general academic skills for advancement into the Junior year requires the choice of a standard or cutoff score. Although much work has been done in the area of standard setting, and although many exams provide such standards, e.g., all licensing exams, no one has developed a very satisfactory method. Usually the cutoff score is heavily influenced by the approximate proportion of examinees who "ought to pass." Furthermore, there is seldom any justification for failing the person one point below the cutoff while passing the person only one point higher. Nor is there an external criterion by which one could demonstrate that those who passed the test have mastered general academic skills, while those who failed have not. One method to soften this reality is to allow students to take the exam or a parallel version more than once. Another frequently used method is to lower the cutoff score to compensate for measurement error. It is important to note that such a practice does not compensate for lack of validity, only for lack of reliability.

Comparisons of results with other institutions should be done

very cautiously, given differences in student entering characteristics, curricula, etc. Obviously when schools choose to use the same national exam and administer it similarly, comparisons may be available. Since many institutions are apparently using the ACT COMP tests, national comparisons are probably available for the various parts of this test. Special problems are associated with comparisons of gain scores, which are covered in the section on value-added testing. However, let us say here that differing entering characteristics may cause students to progress at different rates, independent of the instruction which is received.

G. Sampling. Obviously testing for student placement or advancement into the Junior year requires exhaustive testing of all appropriate students. General academic skills testing for program evaluation and curricular decisions can be accomplished through testing a random sample of students. Depending upon the diversity of the group, a sample of from 250 to 500 may yield sufficiently representative data from which to draw generalizable conclusions. If subsets of students with particular coursework or background characteristics are of special interest, stratified random samples can be tested. In either case, significant cost savings can be realized due to the reduced set of students to be tested.

V. In-Depth Testing

In contrast with general academic skills testing, in-depth testing is expected to cover specific disciplinary areas. Basically, in-depth testing assesses the competence of students in their major fields. These tests would be given at or near the end of a student's undergraduate career. Passing such a test might be one of the criteria for graduation. Certainly, test results would be expected to be a key ingredient in a department's evaluation of its educational program. As above, in-depth testing will be discussed in the context of the seven issues.

A. Test Availability. There are national tests available for some subjects, most notably the Graduate Record Examination (GRE) area tests for some 17 fields such as English, music, sociology, and Spanish. In addition, many of the professional programs can draw upon licensing-type exams. These national exams have the benefit of providing some normative information. They have the disadvantage of possibly not adequately representing the curriculum and goals of a given department. Clearly, such exams need to be chosen deliberately, with careful analysis of what is measured and the meaning of scores.

Tests are not available for many disciplines; thus, if in-depth testing is to be done it will have to rely on locally-developed tests. Given the importance of faculty involvement in the process and of finding a fit with the goals of the program, developing tests locally is not necessarily disadvantageous. It is true that such tests might be criticized for narrowness, for incorporating an inappropriate view of the discipline, or for poor psychometric quality. These criticisms can be partially overcome by having representatives of the discipline from peer institutions share in the process. Also, for many disciplines it would also be reasonable to have appropriate members of the community join in the effort as well, such as potential employers, members of the arts community, etc.

B. Cost. The cost of using standardized tests to examine large numbers of seniors is apt to be high. For example, the GRE area exams cost $29.00 per person to administer. Locally administered tests would represent considerably less out-of-pocket expenditures but a considerable expenditure of faculty time. Presumably members of the community would be willing to volunteer their time, but faculty from campuses in other states would probably expect honoraria, and all off-campus faculty would need travel funds. The cost of faculty time would need to be balanced against the potential benefit of this activity.

C. Reliability. Standardized tests are likely to have excellent reliability whereas the reliability of locally developed tests will have to be determined. The latter may make greater use of essay type questions which can be more difficult to score reliably. Such items should be independently scored by at least two persons and data on scoring reliability should be collected and periodically evaluated. Reliability considerations are more important in the context of testing as a graduation requirement, where decisions affect individuals than for program evaluation. The former requires a higher standard of reliability.

D. Validity. Earlier we mentioned two types of validity, predictive and content. It is unlikely that predictive validity would be a useful concept to employ with in-depth testing, since an acceptable criterion relating to post-graduate success would be hard to develop, given the diversity of what graduates do with their degrees. Correlating test results with grades in major courses would certainly be an interesting exercise, but it is not clear what low correlations would mean, should they be found. Do they reflect a problem with the test, with grading practices, or with the fit between the two?

All things considered, content validity would be the critical validity concern. One reason we suggested that faculty must be intimately involved in choosing or developing tests is to assure the content validity of the result. The outside consultants from peer institutions and from the community would be especially valuable contributors to the content validity of the resulting test. It is our view that validity could be achieved by such a process.

E. Usefulness. In-depth testing of students within their majors could be very useful in evaluating the offerings of each discipline. In the first place it would force faculty to grapple with the issue of just what is expected from someone majoring in their discipline. Secondly, it would present an avenue, beyond course grades, to assess whether those expectations were being met.

The usefulness of testing in order for Seniors to be eligible for graduation depends upon what is done with those who fail. Procedures must be established to enable them to remove their particular deficiencies and re-take a parallel version in order to graduate without undue stress or inconvenience. Otherwise, the process may become a deterrent to advanced education, with people being wary of spending four years of study only to be denied at the last hour. The state would thereby lose valuable education of its citizenry.

F. Standards. Appropriate standards are a problem in using in-depth testing both for evaluating students and for evaluating programs, but more for the former. How can a minimum standard be

established for whether or not a student knows enough of his/her major to graduate? The same reasons why standard setting is problematic given above under general academic testing also pertain here. A further problem is that there is no way to equate those standards across majors? To implement such a policy the institution would have to establish general guidelines, with their interpretation being left to each department. Several years of administering the test for calibration without making actual decisions would be in order.

Standards for program evaluation are equally problematic, but their consequences are not apt to be as disruptive. In most cases departments would have to develop an internal standard against which to evaluate the performance of their majors. Again there could be institution-wide procedures for developing these standards. We see no way to compare departments as a result of the tests, however. Furthermore, comparisons of the same departments across universities is fraught with difficulty, especially if the same exams are not being used. Even with the same exams, curricula differ as do the populations of students from which each school draws.

G. Sampling. Clearly, if graduation decisions are to be made on the bases of in-depth testing, all undergraduates will need to be tested at or near the end of their Senior years. Likewise, for program evaluation, all Seniors will have to be tested for all or almost all majors, because of the relatively small number involved. Where sampling procedures can save resources is in sampling departments on a rotating basis. Perhaps every department can test every five years, or for smaller departments, perhaps every year for two or three years, followed by a five year hiatus. On the other hand, if the test is developed and the number of potential graduates is not large, annual testing may be fully justified.

APPENDIX B

VALUE ADDED TESTING

(Robert M. Thorndike)

A. General.

"Value-added" testing is the assessment of gains in competence that students derive from a college education. Its measurement usually involves giving students a test at some early stage in their college programs and administering a second test at some later date, often at the end of the sophomore year or near the time of graduation. The difference between scores obtained at these two points is attributed to the effect of the college education.

There are three general topic areas in which it may be reasonable to assess changes in students. These topics include cognitive skills, social skills, beliefs and attitudes. It is quite likely that the college experience has a substantial effect in each of these areas. It is also possible to describe in a general way the desired changes in cognitive and social areas. Tests (in the case of cognitive skills) and questionaires (in the case of social skills) can monitor the acquisition of desired changes. In fact, such monitoring is currently done informally at the State's universities. Desirable outcomes are detailed in the curricula of the institutions and achievement of these outcomes is assessed in the process of assigning grades and credit. However, for the purpose of assessing growth, it is necessary to add a pretest to give a point from which to measure gain.

Benefits may result from instituting a program of value-added testing. Benefits include the possible use of test results to evaluate how well each institution is accomplishing its objectives (as set forth by either the institutions themselves or by State policy makers), the evaluation of faculty effectiveness, and the determination of what changes actually take place in students as compared to the assumed changes.

B. Pitfalls

While there is considerable intuitive appeal to the idea of giving a test to students as they enter college and a second test as they leave, such a plan has its problems. It is not appropriate simply to look at the gains in test scores and interpret them as due to the college experience, nor is it appropriate to use gain scores in inter-campus comparisons. The assessment of gains has been one of the most difficult problems in education for years. Some of the difficulties in interpreting the results of value-added studies include:

1. A change that appears to be due to education may in fact be due to living longer. People change continuously and unless

the growth in a group of college students is compared to that of a group of people of similar age and other relevant characteristics who do not attend college there is no way to know, except within defined curricula, that education was an important factor in the change.

2. Different people react to the same "treatment" differently. This problem, known as "aptitude-treatment interaction", has been studied for twenty years. Some students learn better in some environments than others. The students self-select instituitions, so the institution are serving quite different student populations. Findings in one institution may not apply at another, and comparisons between instituions are almost certain to be invalid.

3. All longitudinal studies are affected by nonrandom "subject mortality". As many as half of the students who start a program will not finish at the school where the program was begun; or, they will not finish in the normally-allotted time. In both instances, if these students' pre-test data are included in data analyses, the findings will be affected. Many of those available for post-testing will have transferred into the institution after completing their sophomore year elsewhere, and this factor means that use of their post-test data in the analyses will complicate interpretation of the findings because pre/post groups will not be comparable. What would results be if dropouts had remained?

4. There are serious statistical problems with the interpretation of gains. For purely mathematical reasons, those who start highest in a group tend to show the smallest proportional gains. This influence has nothing to do with the effectiveness of the educational programs. It is an unavoidable statistical artifact that is often incorrectly interpreted.

C. Direct Comparisons

While the concept of value-added testing has some very attractive features, the problems listed above and others make direct comparisons of institutions or programs impossible. There is a proper place for applying gains assessment in higher education, and that place is within programs. The value-added testing moivement has the beneficial outcome of calling upon departments and programs to look carefully at their objectives and how achievement of those objectives is measured. However, attempts to compare programs or institutions can only lead to conclusions of questionable validity.

D. Operational Considerations and Timing

The most straightforward way to assess gains is the pretest-posttest method in which a group of students is given a test at the beginning of their program of study and a second test at the end of the program. It is most common to administer the same test twice and to consider the difference (T2 - T1) to be the

amount of gain (in the case of ability) or change.

For the traditional student aged 18-22, pre-testing of "basic skills" (Liberal Arts) is best accomplished at the time of their entry into the college. The post-test may be applied at either at the completion of the undergraduate liberal arts core or at the conclusion of the college career. Of course, the conclusions that could be drawn would differ depending on when the posttest was administered.

The assessment of gains among nontraditional students involves a different series of problems. Such students present a much more heterogeneous population, so the choice of an appropriate pretest is more difficult. The students will bring to the test a wide range of backgrounds, ages and objectives. They tend to be more serious about study, less impressed with authority and have a more well-developed set of views about self and the world. To the extent that different institutions or programs within an institution tend to attract a different mix of traditional and nontraditional students, it will be difficult to make meaningful comparisons.

E. Advanced Assessments

The assessment of gains at higher levels, either in the major or in a concentrated program of study, is in some ways more complicated. Whereas students start college with some level of mastery of the basic skills that are not discipline dependent, they seldom begin their major with such experience. Therefore, it is seldom fruitful to administer the same test twice because their responses on the first test are unlikely to represent much more than chance.

The typical way that this problem is resolved is to assume that any performance above a chance level must be due to the knowledge and skill gained from the major study and therefore the difference between observed performance and chance is called the gain due to study in the major, but this assumption rests on a weak foundation. It would be more appropriate to administer a low level test in the major at entrance and a higher level test at exit. Unfortunately, the problem of interpreting such results is far from straightforward. Unless the items have been properly scaled on a common scale, the relationship between level of the scores on one test and level on the other is uninterpretable. Such a scaling is simple psychometrically, but presents large problems when applied to a university-wide program.

F. Assessment of Student Attitude

The assessment of changes in student attitudes and beliefs can be accomplished with a pre and post administration of an appropriate questionnaire. Some institutions already collect some information of this kind from their entering freshmen, but the cost of a second collection coupled with the expense of data

analysis and the unclear implications of the findings means that the data seldom are used. It is possible to compare Washington college students with a national sample of participating institutions, but few schools anywhere collect data at the later points in their students' careers.

G. Benefits of Value-Added Testing

On the surface, value-added testing seems to offer several benefits. These include:

1. Educational policy makers could determine how well each institution is doing its job. Institutions that show larger gains in student achievement might be viewed as doing more with their resources than those with smaller gains.

2. The effectiveness of faculty members could be compared in a similar fashion and the dead wood eliminated.

3. The degree to which each institution is accomplishing its stated objectives could be assessed. This would require each institution to specify the objectives of its programs to a degree that is not now done, and the preformance of students in achieving those objectives could be determined.

4. Departmental and collegiate objectives could be assessed in a like manner.

5. An assessment of what changes actually take place, when compared with desired changes, could help faculty to modify curricula to achieve desired outcomes. This is perhaps the best and most attainable benefit that might acrue from value-added testing, but it is an entirely "in house" application. For example, the faculty in psychology at Northeast Missouri State University compared the performance of their students with the national sample for the Graduate Record Exam and determined that they should change their curriculum. The decision to do so implies that the curriculum exemplified by the GRE is what the faculty wanted their students to be learning. A comparison of their students' performances with the norms showed areas of weakness relative to the "GRE curriculum." However, if the faculty had a firmly held belief that some other curricular objectives were more important, such a comparison with the GRE would be inappropriate.

6. It would be possible to identify changes in student attitudes and beliefs.

There are certain types of inferences that cannot be justified without use of a pretest - posttest investigation. In general, changes in students are most easily assessed in this way, but the fact that change is found does not automatically mean that the change is due to the college experience. Conversely, a finding that no change occurred does not mean that the program did not have a desirable effect.

In the first case above, it is not appropriate to attribute change to a "treatment" unless we know that no change or considerably less change would have occurred in the absence of the treatment. Comparison of change in college and non-college and people is complicated by (1) the fact that all people change over time and allocating type or extent of change resulting from college attendance is impossible and (2) that these two groups represent two quite different populations before the treatment starts. There is no way to conduct a study that will identify changes that are clearly attributable to the college experience.

To the extent that the various institutions sample from different applicant pools, there is likewise no way that valid inferences about the effects of attending one institution versus another can be drawn. There is also the problem of what is called the "aptitude - treatment interaction." It has been known for several years that some students respond to one kind of treatment and learn well in one mode while that same approach would not work at all well for other students. Students to a certain extent self-select themselves into one or another treatment. There is no way to tell whether achievements with one group of students in the environment of the school they have selected for themselves would have any beneficial effect on the somewhat different clientele of another school with a different program.

A finding of no change is also uninterpretable unless there is evidence of what would have happened to these students in the absence of the college experience. It is possible that going to college prevented an adverse change. Without comparisons that cannot be performed there is no way to know.

H. Cautions

While their appear to be substantial benefits that can be derived from the use of a value-added approach to institutional assessment, most of the attraction is lost when the problems and restrictions necessary for responsible interpretation are considered. As the noted psychologist Carl Bereiter observed in discussing problems in the interpretation of change, "It is only in relation to such questions that the writer has ever heard colleagues admit to having abandoned major research objectives solely because the statistical problems seemed to be insurmountable." In addition to the problems of effect on the curriculum and of sampling mentioned elsewhere in this report, the measurement of gains poses the following difficulties:

1. As mentioned above, use of a longitudinal design introduces problems of nonrandom subject mortality. The results of a value-added study will be distorted by the fact that some students will not take part in the posttest. Those students who do not take part in the posttest will differ from those who do in systematic but unknown ways, and these differences will vary from institution to institution and program to

program. Comparisons of programs will necessarily include these probably irrelevant unknown factors as part of the "gain" due to one program but missing in another.

2. We have known for 85 years about the problem of regression in the measurement of change. This means that in comparing one individual with another or one institution or program with another the amount of gain is highly likely to show an inverse relationship to starting position for purely statistical reasons. Institutions and programs that score well above average on the pretest are likely to score less well above average on posttest (small gain), and the lowest scoring programs on pretest are likely to have a better second showing (large gain).

3. On a test that is to be applied to a wide range of students, there is the possibility that there will be a floor or ceiling effect. Either poor students will be responding at a chance level on pretest, or good students will be unable to demonstrate the full extent of their accomplishment on posttest. In either case, simple gains will be inaccurate indices of the changes that have taken place.

4. When comparisons of ongoing programs are made over a series of years, large initial gains will almost always be followed by smaller subsequent gains. This occurs simply as a result of the change in the base or standard against which the final performance is compared.

5. When the test content remains the same and ability levels change dramatically, there are sometimes changes in the psychological processes being tapped by the test items. For example, computation problems that are difficult for entering students and test the top of their reasoning powers may, at a later stage, test nothing more than the speed and accuracy of their computation skills. Likewise, the same test may assess different cognitive processes in different students when there are large differences in the levels of student functioning.

6. The units of the scale may be quite different depending on the part of the scale being used. Large gains are relatively easier from a low starting point, so the importance of a single unit of change is greater at higher levels of the scale. This fact interacts with the regression effect to mean that good students must show larger proportional "true" gains in order to appear not to be losing ground relative to lower-scoring students.

7. Gain scores are less reliable than either of the scores on which they are based. This means that small apparent gains may not represent any true gain, and that large apparent gains may represent less true gain than do smaller apparent gains. It also means that a separate

assessment of gain would quite possibly lead to a different conclusion regarding gain than what was actually obtained. This reservation is particularly important for comparisons between individuals or programs.

8. Gain scores present serious problems when compared to norms. Gains of 10 units in one portion of the norm scale may have very different meaning from 10-point gains in other portions of the scale, and the meaning also changes depending on the type of norm scale being used.

9. In any comparative use of gain scores it is important to remember that half of any group is by definition below the median of the group. Half of all professional basketball players are below the median average height of professional basketball players. This is true regardless of the fact that the average is 6 feet 7 inches. It is not possible to escape from this problem unless one has a scale composed in units that are independent of the group being measured. There are ways to combat this problem with test scores in general, but there are not any ways to overcome it with gain scores.

APPPENDIX C

ASSESSMENT OF ALUMNI

(Fred E. Fiedler and Douglas J. Allmon)

How has college attendance affected the individual's life and his/her impact on the community? This question deals with the ultimate outcome of education. Its assessment presents several problems. The first of these is, at what point are the potential effects best assessed? The longer we wait after graduation the greater will be the influence of non-college factors that impinge on the individual's life. On the other hand, many effects of college education may not become apparent until some years after graduation.

A number of investigators have sought to examine the effects of college education on the graduate's subsequent life. Most studies have included an evaluation of his/her life, performance as a productive citizen, and experience as a useful member of society. Life style questions measure the extent to which alumni have achieved employment in fields related to their education, earn appropriate incomes, participate in cultural events, and assume civic responsibility (voting, participating in the political process or civic enhancement panels). Affective questions measure extent to which the alumni express satisfaction with their lives, careers, and educations.

In addition, measures should be considered which indicate the individual's moral development. Ideally, a university graduate and citizen would demonstrate some sophistication in ethical aspects of decision-making which affects others. Brief questionnaires are available for this purpose. Many of these measures are not sufficiently reliable or valid for evaluating one particular individual; they are useful aggregate or average measures in establishing whether college has had the desired effect.

The selection of appropriate control groups presents problems discussed earlier. As a result, we cannot readily compare data from graduates with data from those who did not go to college or who went to private colleges; nor can we compare data drawn from graduates of the various State schools. We can, however, make some judgments about the proportion of graduates who lead productive and successful lives according to prevailing community standards. Useful comparisons can be made with data from peer group schools which conduct alumni surveys. Sample questions asked in surveys of this type are shown below:

Sample questions:

Do you own your own home — Yes — No

Have you voted in at least one of the last elections held in your area since you graduated? — Yes — No

Do you serve on any civic action committee or board such as an advisory board for a day care center, or precinct caucus? Yes No

Have you attended a cultural event, play or concert during the past 90 days? Yes No

Did you read any books for your amusement during the last month? 0, 1, 2, 3, 4, 5, more than 5
(Circle correct answer)

Did you attend any plays or movies during the past month? movies 0, 1, 2, 3, 4+
plays 0, 1, 2, 3, 4+

My job is like a hobby to me. Yes.....No

I am often bored with my job. Yes.....No

I feel that I am happier in my work than most other people. Yes No

I am disappointed that I ever took this job. Yes No

The work I do is meaningful to me. Yes No

On my job, I can't satisfy everybody at the same time. Yes No

My life is satisfying to me. Yes No

Are you currently employed? __ Yes; __ No

Please describe specifically what you do

Do you supervise others? If so, how many? 1-5, 6-10, 11-15, 15+

Does your present job require a college education? __ Yes; __ No

To what extent has your college education enabled you to perform your present job in a more effective manner?

Very much so _____Very little

What is your monthly income? $_____

The basic method of evaluating life style and productivity on post-graduation measures utilizes attitude and opinion survey techniques. A survey of the entire population obviously is expensive but politically less liable to criticism; however, a carefully designed and administered sample will provide highly accurate and generalizable findings at a relatively low cost (see Table 4).

A key concern is determining at what point(s) the questionnaires are to be administered. The two most reasonable points appear to be two or five years after graduation. After two years, the individual's life is still directly affected by experiences in college. This influence applies especially to students who entered professional and graduate programs. However, even for those who entered employment, the college experience continues to exert a major influence on the individual's employment and private life.

The methodology of survey research is well known, and the surveys can be conducted under contract by a private organization, under direct supervison of the HEC Board, or by each of the universities and colleges. The actual size of the sample need not be large, unless specific subsamples of students (for example, by department or college) are to be evaluated. Table 4 shows the approximate number of graduates and suggested sample sizes.

Table 4

Approximate Number of Graduating Seniors and Suggested Sample Sizes

School	Graduates 1984-85	Minimum Sample
UW	4,977	400
WSU	2,761	200
CWSU	1,470	200
EWSU	1,439	200
Evergreen	524	200
WWSU	1,737	200
TOTALS	12,936	1,300

Specific Procedures

A possible data gathering method is here outlined:

1. All graduating students are required to provide three names and telephone numbers of relatives who are likely to know the graduate's whereabouts during the next five years.

2. Each four-year institution (or designated programs) provides a list of all students who have obtained bachelors degrees during the year.

3. Two years after graduation, an organization designated by the HEC Board draws a random sample of 400 names from among the graduates of the University of Washington and a sample of 200 for each of the other schools. (Surveys of this type are conducted on a contract basis by NCHEMS, ETS, and similar organizations.

4. After an introductory letter explaining the purpose of the survey, the respondent is contacted by telephone and interviewed for about 30 minutes. (In case of "no answer", typical procedure might require three re-calls, followed by substitution of the next name on the list.)

5. The interviewer administers a short questionnaire by telephone consisting of:

a. Job satisfaction scale: How well does respondent like the work itself, fellow employees, pay, supervisor, working environment.

b. Life satisfaction scale. How satisfied is respondent with his or her present life, place of residence,

available money for necessities and luxuries, health services; expectations for future,

c. Participation in community and political affairs

d. Participation in cultural activities

e. Occupational status: employed in what capacity and what type of organization; self-employed, unemployed, etc.) ETS has scales of this nature.

f. Biographical information: Life style; married, single, divorced, widoved, separated; children; living alone, with others, with parents, with own family); living in Washington or other states; urban or rural; ethnic background, parental mother tongue, parental occupational status.

g. In retrospect, to what extent does respondent feel that college education contributed to present way of living, job performance, and expectations for the future?

6. Analyses will identify respondents by school, by geographical area, by major field of concentration, sex, age, and other relevant variables.

Two options for conducting surveys of alumni are here outlined. To ensure adequate representation of each academic discipline periodic ongoing repetitions of the measurements may be instituted using stratified random sampling of all academic departments: stratification is a method by which a pre-determined number of alumni would be designated in each department for random selection.

Option I

This option calls for polling alumni 12 or 24 months following their graduation. Under this Option, telephone contact will be made to elicit agreement to complete questionnaires which will be mailed; a fee of up to $40 might be offered for timely return of each questionnaire. Expenses associated solely with Option I (these exclude general project expenses such as office facilities, supplies, operating expense, salaries of supervisors, and computer analysis):

Telephone Representatives	$ 6500
Telephone Tolls	2600
Postage (two-way)	2600
Photocopying Questionnaires	2000
Fees to Respondents	104000

Copyrighted Questionnaires and pre-;aid scoring	13000
18 - 24 month Total	$130,700

Option II

Option II replicates Option I and adds a repetition of the first assessment at 60 months. The projected expense estimate reflects allovance for tracing respondents and the influence of inflation.

Telephone Representatives	$ 14,325
Postage (Tvo-vay)	3,900
Photocopying	3,000
Fees to Respondents	104,000
Copyrighted Questionnaires	19,500
60 Month Total	144,725
18 - 24 Month Total	230,700
Grand Total	375,425

Discsussion of Option I and II. Option I has the advantage of producing data which will permit comparisons to findings emerging in similar assessment projects in other states, vhere assessments of alumni satisfaction and affective states are common. All measures of knowledge, grades, change (gain), expertise contemplated in prior sections of this paper are carried out on the assumption that they will measure prospective contribution by alumni in the "real world". The assumption is that knowledge acquired by alumni will be translated into (1) societal gains (economic, cultural, life quality) and (2) enhanced personal experience of life.

Options I and II share the joint disadvantages of adding expense and time to the overall assessment. Their advantage is that they measure actual events, not variables merely assumed to contribute favorably to community well-being and individual quality of life; indeed, Options I and II are the only direct measures of these concerns. Mail surveys have been used by various organizations, but response rates are usually too lov (40-50%). A questionnaire developed by NCHEMS is attached.

Finally, mention should be made of evaluating the alumni of a school by surveys of the State's major employers. The difficulty and cost-effectiveness of this approach is difficult to determine without knoving the number of employers who are to

be contacted. However, it is certainly possible to contact major employers in this state to determine their satisfaction with the alumni from various schools, and to obtain from employers perceptions of the strengths and weaknesses of graduates from various schools.

Summary. Post-graduation assessment of demographic and affective features of the lives of university alumni is becoming more common in examining effectiveness of universities. For purposes of full comparability of data, evaluations of this type need to be made on an ongoing basis and will provide useful information about the contributions of the college education to the student's later life as well as to the life of the community.

APPENDIX D

SAMPLE QUESTIONS

UNIVERSITY

School of Health Related Professions and Natural Science

Alumni Questionnaire

NCHEMS Management Services, Inc.
P.O. Drawer P Boulder, CO 80302

INSTRUCTIONS:

Specific instructions are given for completing many of the questions in this questionnaire. Where no directions are given, please circle the number of the most appropriate response, such as in the sample question below.

Sample:

4. Are you currently married?
   1 Yes
   2 No

If you are not currently married, you would circle the number (2)

PERSONAL DATA AND GOALS

1. What is your sex?
   1 Female
   2 Male

2. How do you describe yourself? Circle one.
   1 Black or Afro-American
   2 White or Caucasian
   3 Other_____

3. What year were you born?
   Please enter the year in the box below.

   [1][9][ ][ ]

4. When you attended
   were you a foreign student?
   1 Yes
   2 No

5. What is the highest level of education obtained by your parents?
   (Circle the appropriate number in each column).

Father	Mother	
1	1	Not a High School Graduate
2	2	High School Graduate
3	3	Some College
4	4	Bachelor's Degree or Above

6. Circle all the degrees you have received from
   University
   1 Certificate
   2 Associate Degree
   3 Bachelor's Degree
   4 Master's Degree

7. In the box below, enter the year that you received your most recent degree from University.

   [1][9][ ][ ]

8. While at which of the following best described your enrollment status?
   1 Full Time
   2 Part Time

9. When did you take the majority of your courses at
   1 Day
   2 Evening

10. What was your typical employment status while attending
    1 Employed 1-9 hours/week
    2 Employed 10-19 hours/week
    3 Employed 20-29 hours/week
    4 Employed 30 or more hours/week
    5 Not employed

11. When you first entered did you (Circle only one)?

1 Enter directly after high school
2 Enter after working (other than a summer job)
3 Transfer from a community college
4 Enter from military service
5 Transfer from a 4-year institution

12. The following statements reflect some of the reasons why people attend college. Circle the numbers that are appropriate in both columns.

These are goals I achieved while I attended

These are goals I have achieved or am achieving since I left

Academic Goals

1 1 An increase in my knowledge and understanding in an academic field
2 2 A certificate or degree
3 3 Completed the courses necessary to transfer to another school

Career-Preparation Goals

4 4 Discovered career interests
5 5 Formulated long-term career plans and/or goals
6 6 Prepared for a new career

Job- or Career-Improvement Outcomes

7 7 Improved knowledge, technical skills, and/or competencies for job or career
8 8 Increased chances for a raise and/or promotion

These are goals I achieved while I attended

These are goals that I have achieved or am achieving since I left

Personal-Development and Enrichment Outcomes

9 9 Increased self-confidence
10 10 Improved leadership skills
11 11 Improved ability to get along with others
12 12 Learned knowledge and skills that will enrich my daily life
13 13 Developed ability to be independent, self-reliant, and adaptable

13. From the list in question 12, please select the three that now seem to you to be the most important results of having attended our college. For example, if you now think that the most important result for you was "improved leadership skills" enter the number 10 in the first box.

Most Important ☐ Second Most Important ☐ Third Most Important ☐

14. Indicate how helpful your experiences at were in achieving the following:

1 Very helpful
2 Helpful
3 Somewhat helpful
4 Not helpful
5 Does not apply to me

1 2 3 4 5 Learning how to think and act independently
1 2 3 4 5 Writing papers
1 2 3 4 5 Putting ideas together from different sources
1 2 3 4 5 Learning that there are many different points of view
1 2 3 4 5 Developing interest and appreciation of the arts
1 2 3 4 5 Using mathematics and quantitative methods
1 2 3 4 5 Tolerance and understanding of other people

1 Very helpful
2 Helpful
3 Somewhat helpful
4 Not helpful
5 Does not apply to me

1 2 3 4 5	Appreciation of individuality and independence of thought and action
1 2 3 4 5	Awareness of different philosophies, cultures, and ways of life
1 2 3 4 5	Learning how to think critically
1 2 3 4 5	Gaining skills in computer usage
1 2 3 4 5	Making oral presentations
1 2 3 4 5	Developing interest and appreciation of Afro-American experiences
1 2 3 4 5	Developing a set of values for my personal and civic life

CAREER/EMPLOYMENT

15. Since you completed your program at . have you been employed for pay?
    1 Yes, I am currently employed full-time
    2 Yes, I am currently employed part-time
    3 Yes, I was employed after leaving Norfolk State but currently am unemployed and looking for work
    4 Yes, I was employed after leaving Norfolk State but currently am unemployed and not looking for work
    5 No, I was never employed after leaving Norfolk State (If No, skip to question 21)

16. a. What was the first job you had after you completed your program at (e.g., accountant, engineer, salesperson, teacher)?

_____

b. What is your current job? If you are not currently employed, what was your most recent job?

_____

17. Circle the number in the first column which best describes your salary in the first job you had after completing your program at
Circle the number in the second column which best describes your current salary.

Starting salary in your first job after college	Salary at current job
1	1 Less than $12,000 per year
2	2 $12,000 to $17,999 per year
3	3 $18,000 to $23,999 per year
4	4 $24,000 to $29,999 per year
5	5 $30,000 to $35,999 per year
6	6 $36,000 to $41,999 per year
7	7 $42,000 per year or more
8	8 Does not apply

18. Is your current job related to your major field of study at ?
    1 Directly related
    2 Somewhat related
    3 Not related at all

19. How well did prepare you for the work you are now doing (or if you are not currently employed, for the work you did most recently)?
    1 Excellent preparation
    2 Good preparation
    3 Adequate preparation
    4 Inadequate preparation

20. Which statement best describes how you regard your current job? (Circle one)
    1 Temporary job
    2 Job with possible career potential
    3 Job with definite career potential
    4 Other (Please specify) _____

21. Since completing your program at State, have you undertaken further formal study?
   1 Yes
   2 No (If No, skip to question 27)

22. a. After your program at what was the institution that you most recently attended:
   1 A public two-year college
   2 A public four-year college, university or professional school
   3 A private college, university or professional school
   4 A vocational/technical school, trade school, or business school
   5 Other (specify)_____

   b. Please write in the name of that institution.

_____

23. What degree or certificate were you seeking at that institution?
   1 Certificate
   2 Associate Degree
   3 Bachelor's Degree
   4 Master's Degree
   5 Specialist or Doctoral Degree

24. Have you completed the requirements of this certificate or degree
   1 Yes
   2 No, but I am still enrolled seeking it
   3 No, but I still plan to complete it even though I am not currently enrolled
   4 No, and I do not expect to complete it
   5 Does not apply

25. Please write in the major or area of study associated with this degree or certificate.

_____

26. How well did _____ prepare you for your additional formal education?
   1 Excellent preparation
   2 Good preparation
   3 Adequate preparation
   4 Inadequate preparation

27. During the past year, about how often have you done the following:
   1 Not at all
   2 Once or Twice
   3 About once/month
   4 About once/week

1 2 3 4 Contributed time to a civic or neighborhood project or institution, such as a playground, school, hospital, etc.
1 2 3 4 Gave money to a community fund, community chest or other local charity
1 2 3 4 Had contact with a local offic about some problem
1 2 3 4 Attended meetings of a politic club or group
1 2 3 4 Visited an art gallery, art mu etc.
1 2 3 4 Attended one or more musical performances

EXPERIENCES AT _____

28. We would like to find out how satisfied you were with various aspects of academi life at _____ University.
   1 Very Satisfied
   2 Satisfied
   3 Somewhat Satisfied
   4 Dissatisfied
   5 Very Dissatisfied

1 2 3 4 5 University General Education Requirements
1 2 3 4 5 Variety of courses offered in my major area of study
1 2 3 4 5 Adequacy of instructional equipment in my major departmen
1 2 3 4 5 Quality of my internship, practicum or co-op if applicabl
1 2 3 4 5 Interest of the faculty in students
1 2 3 4 5 Quality of instruction
1 2 3 4 5 Faculty's interest in teaching
1 2 3 4 5 Availability of faculty advisor
1 2 3 4 5 Quality of faculty advisement

29. Please indicate your satisfaction or dissatisfaction with the following services or activities by circling the appropriate number. If you did not know about these services, please put a circle around the number 5 under the column "Did Not Know About". If you knew about the services, but did not use them, circle the number 6.

1 Very Satisfied
2 Satisfied
3 Dissatisfied
4 Very Dissatisfied
5 Did Not Know About
6 Did not Use

1	2	3	4	5	6	Admissions process
1	2	3	4	5	6	Registration procedures
1	2	3	4	5	6	Business Office
1	2	3	4	5	6	Counseling Center
1	2	3	4	5	6	Testing
1	2	3	4	5	6	Financial aid office
1	2	3	4	5	6	Tutoring
1	2	3	4	5	6	University cultural programs
1	2	3	4	5	6	Recreation and athletic programs
1	2	3	4	5	6	Job/career placement office
1	2	3	4	5	6	Career Day
1	2	3	4	5	6	Health services
1	2	3	4	5	6	Library services
1	2	3	4	5	6	Cafeteria
1	2	3	4	5	6	Bookstore
1	2	3	4	5	6	Campus security
1	2	3	4	5	6	NSU Housing
1	2	3	4	5	6	Computer Services
1	2	3	4	5	6	Others_____
1	2	3	4	5	6	_____

30.

31. What is unique about Uni

_____

_____

_____

_____

TEACHER EDUCATION

If you were in teacher education (secondary or elementary) while at University, please answer questions 32-38. If not, please skip to the next section (question #39).

32. Please indicate how satisfied you were with the following experiences while participating in the teacher education program.

1 Very Satisfied
2 Satisfied
3 Indifferent
4 Dissatisfied
5 Very Dissatisfied

1	2	3	4	5	Your professional education requirements
1	2	3	4	5	The dual advising program (Major Department/School of Education)
1	2	3	4	5	Your student teaching experience
1	2	3	4	5	The evaluation procedures for student teaching
1	2	3	4	5	The assistance of the university in helping you get your first job
1	2	3	4	5	Overall teaching preparation (at 's School of Education)

33. Are you currently employed in the field of education?
    1 Yes (skip to question 35)
    2 No

34. If no, what was the reason you decided to seek employment outside of teaching? (Circle the appropriate number) (Skip to question 39)
    1 Did not wish to teach
    2 Unable to find employment
    3 Geographically unsuitable
    4 Present job more lucrative
    5 Better advancement opportunities in present job
    6 Loss of interest in teaching
    7 Other (specify) _____

35. How satisfied are you with the following aspects of your current job?

    1 Very Satisfied
    2 Satisfied
    3 Indifferent
    4 Dissatisfied
    5 Very Dissatisfied

    1 2 3 4 5 Achievement
    1 2 3 4 5 Job Security
    1 2 3 4 5 Recognition
    1 2 3 4 5 Salary
    1 2 3 4 5 Personal Growth
    1 2 3 4 5 Professional Growth
    1 2 3 4 5 Advancement
    1 2 3 4 5 Professional Autonomy
    1 2 3 4 5 Overall job satisfaction

36. How satisfied do you think your employer is with your job performance?
    1 Very Satisfied
    2 Satisfied
    3 Indifferent
    4 Dissatisfied
    5 Very Dissatisfied

37. What is your long-term professional goal in education?
    1 Teacher
    2 Supervisor
    3 Principal
    4 Superintendent of school district
    5 Educational researcher or writer
    6 University professor
    7 Other _____

38. If you do not plan to make teaching your life-long career, list the areas that you are currently considering.

    _____

    _____

    _____

SCHOOL OF HEALTH RELATED PROFESSIONS AND NATURAL SCIENCES

39. What was your departmental major in the School of Health Professions and Natural Sciences?

_____

40. How satisfied were you with the efforts of your Department in preparing you for a specific field of employment or career? (Circle the appropriate number)

1 Very Satisfied
2 Satisfied
3 Somewhat Satisfied
4 Not Satisfied
5 Does not apply to me

1	2	3	4	5	Opportunities to talk and visit with people who are successful at what I wanted to do
1	2	3	4	5	Educational facilities and equipment that is appropriate for my training
1	2	3	4	5	The specific content of courses in my major field
1	2	3	4	5	Skills and techniques directly applicable to my career
1	2	3	4	5	Overall preparation for my career

41. Have you had a full time job related to your field of study?
1 Yes
2 No (If no, skip to question 47)

42. Were you expected to take a licensing or certification examination in your field of study?
1 No, and I do not plan to
2 No, have not taken an exam but plan to
3 Yes, but did not pass
4 Yes, but don't know results
5 Yes, and I passed exam

43. How important is it for your program of study to be accredited?
1 Not important
2 Important
3 Very important
4 Don't know

44. Indicate the degree of influence your school had on broadening your understanding of the moral and ethical issues facing your profession.
1 Does not apply to me
2 No influence
3 A little influence
4 Considerable influence
5 A great deal of influence

45. Please list the three courses that you took at University that were most relevant to your current job (including internships).

_____
_____
_____

46. In what areas of knowledge and skills do you think you needed more training? List up to three areas starting with the most important need area.

_____
_____
_____

47. If you are currently not employed or are not working in a job related to your major area of study at indicate if the following are true or not true.

1 Not true
2 True

1	2	I never looked for work related to my field of study
1	2	I looked for work related to my field of study but could not find any

48. If employed in a field other than your training area, please state your current occupation:

_____

APPENDIX E

EXPENSE CONSIDERATIONS

TABLE 1

Case 1—Private Liberal Arts College

Instrument Costs	
300 Freshman General Education Exams (ACT-COMP)	$4,500.00
150 Senior General Education Exams (ACT-COMP)	2,250.00
150 Senior Activity Inventories (ACT-COMP)	525.00
300 Freshman Surveys (CIRP)	415.00
150 Current Student Surveys (Pace CSEQ)	337.50
150 Alumni Surveys (ACT-ESS)	147.50
Administration Costs	
In-Class Test Administration	
Proctors, etc.	
Announcements, etc.	342.00
Mailed Survey Costs (2 mailings)	177.50
Overhead/Analysis Costs	193.62
ACT Comp Consulting Visit (Fee + Travel)	
CIRP Data Analysis	1,375.00
Testing/Measurement Specialist (.35 FTE)	150.00
Secretary/Clerk (.35 FTE)	9,625.00
Staff Benefits	5,775.00
TOTAL	3,388.00
	$29,201.12

Case 2—Major Public Research University

The second case is a major public research university with a total enrollment of over 25,000 students, including about 18,000 undergraduates. Faculty make considerable research contributions to their own disciplines and concentrate much of their teaching energy on graduate instruction. Most introductory courses are lecture classes and are partially staffed by graduate teaching assistants. Most undergraduate students attend full-time, and about two-thirds are residential. Attrition rates are significant, but about 65% of entering students complete their degrees. Professional schools account for approximately 60% of undergraduate enrollment.

Because of its emphasis on professional and pre-professional study, much assessment effort has gone into testing in the major field. Graduates of about 10 programs per year are tested using available standardized test instruments. This year, 450 students are to be tested using a variety of GRE Field Examinations, and 360 students are to be tested using pre-professional examinations such as the National Teacher Examination (NTE) and the AICPA exam. In addition, the institution is evaluating general education using the ACT-COMP Objecuve Test in a test-retest format for freshmen and seniors. Like Case 1, the institution has budgeted for a consulting/faculty development visit in conjunction with the COMP.

To examine student life, the university intends to design its own survey, using faculty expertise. The survey will be administered to a stratified random sample of currently enrolled students in the spring. Because of the size of the campus and the characteristics of the sample, a mailed format will be used to administer the survey. Similar surveys in the past have obtained a 65% response rate.

To coordinate the testing program, the university has staffed an existing student research office with two new staff members, a testing specialist and a secretary. Existing senior staff in the testing office are also used in interpreting test results and in working with individual program faculties on improving curricula. Because

TABLE 2

Case 2—Major Public Research University

Instrument Costs	
2500 Freshman General Education Exams (ACT-COMP Objecuve Test)	$15.000.00
1700 Senior General Education Exams (ACT-COMP Objecuve Test)	10.200.00
450 Senior Field Exams (GRE)	13.050.00
360 Senior Field Exams	7.500.00
Development Cost for 3 Local Field Examinations (Professional and Pre-Professional)	9.270.00
Development Cost for Student Survey	5.200.00
2025 Surveys (Production and Scoring Cost)	518.75
Administration Costs	
In-Class Test Administration	
Proctors, etc.	1.826.00
Announcements, etc.	2.077.00
Mailed Survey Costs (2 mailings)	1.957.00
Overhead/Analysis Costs	
ACT Comp Consulting Visit	1.375.00
ACT-COMP Data Tape	20.00
Testing/Measurement Specialist (1 FTE)	27.500.00
Secretary	16.500.00
Staff Benefits	9.680.00
Office Expenses	8.400.00
TOTAL	$130.073.75

TABLE 3

Case 3—Regional Comprehensive University

Instrument Costs	
300 Freshman Tests (ACT Assessment: 900 assumed to have scores on entrance)	$2,100.00
800 Sophomore Tests (ACT Assessment)	5,600.00
1200 Freshman Interest Inventories (ACT Assessment)	3,000.00
80 Senior Field Exams (GRE)	2,320.00
40 Senior Field Exams (Professional and Pre-Professional)	940.00
1200 Entering Student Surveys (ACT-ESS)	240.00
350 Non-Returning Student Surveys (ACT-ESS)	70.00
650 Alumni Surveys (ACT-ESS)	130.00
Scoring for 2200 ACT-ESS Instruments	1,040.00
Administration Costs	
In-Class Test/Survey Administration Proctors, etc.	
Announcements, etc.	375.00
Mailed Survey Administration (2 mailings)	580.00
Overhead/Analysis Costs	1,378.00
ACT-ESS Tape/Reports	
Testing/Measurement Specialist (.35 FTE)	270.00
Staff Benefits	9,625.00
Work Study Students	2,118.00
Office Expenses	1,750.00
TOTAL	1,250.00
	$32,786.00

TABLE 4

Case 4—Mid-Sized Community College

Instrument Costs	
750 Sophomore General Skills Exams (CLEP General)	$19,500.00
4500 Entering Student Surveys (SOIS)	675.00
1500 Former Student Surveys (SOIS)	225.00
1250 Graduate Follow-Up Surveys (SOIS)	187.50
Scoring for 7250 SOIS Instruments	3,490.00
Administration Costs	
In-Class Test Administration Proctors, etc.	
Announcements, etc.	325.00
Mailed Survey Administration (2 mailings)	225.00
Overhead/Analysis Costs	3,693.00
Tapes/Reports of SOIS Surveys	
Student Survey Coordinator (.5 FTE)	150.00
Staff Benefits	10,750.00
Office Expenses	2,265.00
	1,500.00

APPENDIX F

REVIEWER COMMENTS

NCHEMS
National Center for Higher Education Management Systems
P.O. Drawer P/Boulder, Colorado 80302/(303) 497-0301
An Affirmative Action/Equal Opportunity Employer

DENNIS JONES
President

BEN LAWRENCE
President Emeritus

BOARD OF DIRECTORS

SHERRY H. PENNEY
(Chairman)
SUNY–Central Administration

KENNETH P. MORTIMER
(Chairman Elect)
The Pennsylvania State University

ARTHUR G. ANDERSON
IBM Corporation

ROBERT G. BONHAM
Cities Rubber Co., Retired

WILLIAM E. DAVIS
Oregon State System of Higher Education

NORMA FOREMAN GLASGOW
Connecticut Board of Governors

KERMIT HANSEN
Financial Perspectives Company

DAVID P. JOHNSON
National Institutes of Health

JOE B. RUSHING
Tarrant County Junior College

JAMES I. SPAINHOWER
Lindenwood College

WILLIAM E. VANDAMENT
The California State University

HON. GORDON O. VOSS
Minnesota State Representative

MARVIN WACHMAN
Temple University

K.B. WATSON
Pitney Corporation, Retired

EX-OFFICIO MEMBER

JAMES A. HYATT
Chairman
National Advisory Council
National Association of College and University Business Officers

To: Prof. Fred E. Fiedler, University of Washington

From: Peter T. Ewell, NCHEMS PTE

Re: Report to HEC on Evaluation and Assessment

Date: January 1, 1987

As requested, I have reviewed the document prepared by your committee on the state of the art in assessment practice and recommendations for assessment policy for Washington higher education, to be submitted to the state Higher Education Coordinating Board. As I mentioned on the phone, I believe that the document does what it set out to do, and I have no major suggested revisions or amendments. The appendices on current practice in testing, "value-added" research, and alumni follow-up are accurate and are particularly helpful; the substantive recommendations for board policy presented in the body of the document are appropriate and doable. Given the short time allowed to undertake this task, the committee has accomplished much.

The intent of this memo is to elaborate on our phone discussion and to raise a few additional points for your consideration. Most of my reactions are best presented within the context of one or more of your committee recommendations, so I will present muy comments in this fashion.

1. The assessment of performance of educational institutions is indeed "needed, feasible, and difficult." Sensitivity for these apparently conflicting attributes is evident throughout the report. But I was particularly gratified to see attention paid to the "consumer" issue -- who exactly is this evaluation for? This is the single most important question that any state must face, and different states have to date answered it quite differently. Florida, for example, sees the ceritification of individual student abilities as important and has designed its system accordingly. Tennessee and others have concentrated on particular dimensions of performance with the primary consumer the higher education system itself. How the question is answered will, of course, profoundly influence design.

In our work with various state governments, we generally think of the "consumer" question as two questions -- (a) is the intent of the evaluation primarily to provide accountability or to improve performance, and (b) what is the primary unit of analysis for evaluation. Combinations of answers entail quite different designs. Furthermore, systems designed for one purpose (eg. "rising junior" exam programs) rarely provide much of a basis for attaining additional objectives.

2. Diversity among institutions indeed demands multiple measures and a variety of evaluation approaches. One additional point to emphasize is that state-level evaluation questions may substantially differ from those raised at the institutional level. State-level questions are complex because the state needs to know both (a) whether individual institutions are in some sense "effective", and (b) how the entire system of higher education is furthering state objectives. While some common measures are appropriate to address the former question, most such information is best gained by requiring or encouraging a range of institutional self-studies. The latter question may require independent methods addressed around the primary reasons why public higher education is funded. Traditionally, these reasons include (a) ensuring relative equity in the distribution of higher education's (presumed) benefits by removing barriers to access, (b) promoting economic development through appropriate research and training to meet identifiable manpower needs, and (c) ensuring a minimum level of competence or "functional citizenship" among the inhabitants of the state.

3. I fully concur with your discussion on tests for general education. Note however that for many students "general education" does not end with the sophomore year --either because core or distribution requirements are not yet met or because (as most curriculum designers contend) the major continues to teach "general" skills as well. Many emerging assessment programs therefore examine general education at the end of the senior year rather than at the end of two years.

I also support the notion that "general education" assessment should not be about basic skills but should instead concentrate on such "higher order" attributes as critical thinking and problem solving. Note that although ACT-COMP is presently the leading seller in this arena, other commercial alternatives are emerging. You might particularly watch for an ETS entry in the coming year that will rely on item banks rather than a single instrument. Given its flexibility, such an approach might be particularly appropriate to a state system.

4. Your discussion of existing major field examinations is a good one. In our experience, existing standardized exams will cover about half the fields offered by any given institution. At the same time, we have found that faculty-designed local exams can be particularly beneficial to curriculum development and therefore tend to encourage their use. We also like to encourage faculty to think beyond "paper-and-pencil" tests alone. Peer evaluators or visiting examiners are excellent examples in this area, as are auditions in the performing arts. Some language to the effect that this is OK might avoid the common tendency to dismiss non-standardized approaches. Finally, while sampling is an excellent cost-saver, as you point out, it raises some questions about student motivation since individual student fates do not depend upon test performance. Most institutions end up using samples anyway, but the objection can be a real one.

5. Your discussion of the limits of "value added" measurement approaches is particularly valuable. In general, most such approaches, if taken literally, are badly flawed. Certainly the "metaphor" of value-added is a powerful one and all assessment should be developmental and longitudinal. But this by no means demands test-retest on the same instrument. As an alternative, it is quite possible to use entrance or placement test scores as statistical controls in interpreting later achievement results -- just as it is appropriate to use demographic and "treatment" factors. I would strongly encourage making such entrance data part of the general assessment database without assuming that all "gain" is a product of treatment. As you point out, simple difference scores are fraught with peril.

Your cautions about assessing attitude/value development are well taken. Note, however, that there are some values that might have more general currency as desired "outputs" of the college experience. These include integrity, responsibility, persistence, relative value neutrality, and potential for leadership. Some of these may be particularly important to particular "technical" fields, for example Nursing, Accounting, etc.

6. Alumni contributions to society, work, etc. are especially important questions because they represent a major reason for public investment in higher education. For reasons of practicality, however, most state agencies do not undertake alumni follow-up studies directly, but rather encourage institutions themselves to periodically conduct such studies. Given considerable diversity among institutions, this might be the best approach for Washington

as well. In order to ensure some commonality, the state might mandate inclusion of ten or so common items on such surveys, leaving the institutions free to ask any other questions they like.

As I mentioned on the phone, sample size issues can be politically complex. If a survey is administered centrally for limited purposes, small samples will suffice. Almost inevitably, however, the results of surveys become controversial if comparisons among institutions or programs are made. The results of sample studies then tend to fare badly. Also if the institutions themselves are to derive any benefit from the survey in terms of their own planning and evaluation efforts, samples will need to be larger than those suggested.

Finally, note that many benefits of higher education are potentially experienced by non-graduates as well as "alumni". Any potential survey design should include former students as well as those officially graduated.

7. I completely concur with your cautions about directly comparing outcomes results from different types of institutions. It is for this reason that we encourage the use of carefully selected peer institutions in most of our work. Certainly one benefit of using readily available assessment instruments is the probability that some of these peers have comparative data available.

Again, however, I would stress the role of the HEC in encouraging institutional self-assessment rather than emphasizing a common database for comparison. Self-study guidelines can be structured in such a way that there are a limited number of common data elements requested, but that most common questions can be answered in appropriately different ways by different institutions.

8. Faculty involvement is indeed critical, and it is important to call attention to it. Most institutional assessment programs begin with a faculty committee on each campus that spends up to a year examining existing models, instruments, and approaches before recommending campus policy. Your recommendation rightly highlights the need for such an approach.

9. Consistent with your final recommendation, it is in our experience critical to allow sufficient time for a good approach to be developed. Too many states are currently embarking upon "crash programs" whose results may be questionable. Three years is not too long to allow a program of quality to emerge. Your recommendation for a pilot

program or the use of "demonstration campuses" is also a good one. This approach is currently being used in several states and has proven beneficial.

All told, the report should provide a firm point of departure for the HEC in evolving state policy on assessment. Thanks again for the opportunity to review the document, and good luck in its future development.

Memo re draft report for the Higher Education Board of Washington:

This is a very sophisticated report, and my comments are directed more toward highlighting some of the issues discussed rather than to suggest major modifications. In general I will follow the report page by page.

In the Executive Summary I would expand the word"knowledge" to indicate that you are referring to knowledge in the broader sense of knowledge of skills and strategies as well as simple knowledge of facts. What we hope our students achieve is facts organized in a framework of theoretical concepts which can form the basis for further learning and thinking--not simply isolated facts such as are tested by true-false or most teacher-made objective tests. Moreover we are trying to help students develop skills and strategies for further learning as well as for problem solving and decision making. When we set out to measure "core knowledge" we all too often think of dates like 1776 or names like DesCartes or definitions of terms rather than the methods of learning and thinking that characterize our disciplines and that can be transferred to other situations requiring learning and thinking.

Although discussed later, I was struck by the omission of motivational outcomes in the executive summary. Surely one of the most important goals of higher education is to develop an eagerness to learn--curiosity about the meaning of one's experiences, sensitivity to the complexities of life, a favorable attitude toward continued learning in one's vocation as well as in life in general.

I would make Point 7 in the executive summary even stronger. Attempts to compare different institutions always involve comparing apples with oranges. No two institutions can be evaluated on the same dimensions because their students, their milieu, as well as their faculty, curriculum and goals differ. It is true that we can discriminate a good apple from a bad orange, but we should be wary of assuming that scores on tests or questionnaires mean the same thing at different institutions. Your analysis in the appendix of the problems with Sandy Astin's "value-added" idea is well-done. Using measures of student achievement as measures of institutional effectiveness is always more complicated than is apparent on the surface. Even if we compare students of the same ability level at the institutions compared we cannot match them on the basis of the motivational and attitudinal variables that resulted in the choice of different institutions for matriculation.

Obviously the Ann Arbor News takes pleasure in pointing out the high achievement of the students in Ann Arbor schools on the state-wide assessment tests, and even though commentators point out that Ann Arbor

has a high proportion of families with PhD. parents, there is always the implication that the Ann Arbor schools must be doing a better job than Detroit inner-city schools whose students come out with lower scores.

Similarly it would not make sense to compare Michigan and Michigan State, because we have different images in potential students' minds and even students of comparable backgrounds select one or the other because of different expectations about the amount of time they will study, the degree of involvement with social life, and the vocational or cultural goals that may be achieved by matriculation. Even comparisons between Eastern Michigan, Western Michigan and Central Michigan are dubious because of the differences in the high school background and family background of the regions from which their students tend to be drawn.

I also applaud points 8 and 9 about the importance of faculty involvement and of pre-testing what program is devised.

On page 5 "Testing" I would once again emphasize under "liberal education" general skills in learning and problem solving.

On page 6 I liked your seven issues to be addressed in testing.

On page 7 I would underline your caution about requiring that seniors pass the test in order to graduate. In addition to the problems you adduce I would stress the problem of the stress this imposes on the students. We have shown that some students with a good conceptual structure at the end of a course do poorly on the final examination because of high levels of test anxiety. Your own research demonstrates this in another setting.

On page 8 and 9, there may be an implication that self-report measures of attitudes and values should be avoided because of differences in goals. I think that there is more consensus that you imply. As I've already mentioned I believe that most people would agree that a positive attitude toward learning is a good thing, and even in areas in which there may be disagreement about the particular solution to an ethical problem, we might still find agreement that students should become aware that an ethical problem is involved. I think many of the problems in our society are caused by individuals who undertook actions with unfortunate consequences without thinking of the possibility that an ethical issue might be involved.

On page 9 I would add to your measurement methods that self-report measures may also be appropriate for assessing behavior--not necessarily perfectly valid, but with enough validity to be useful, for example, in assessing methods of studying, participation in cultural activities, etc. I liked what you suggest with respect to alumni questionnaires.

I have a note on Table 3 which is answered later, but I'll mention it anyway. It is, "What about ATI's?" I think assessment may be particularly useful if the data are examined to look at the success of a program or an institution in educating differing types of students. I think it is a mistake to look at only the means, or at the very bright, who go on to graduate work. We need to look at the achievement of those who enter with some deficiencies in background, or at minorities, or at rural vs. urban students, etc. to see if we're missing some groups of students.

I certainly agree that one does not want an assessment system which homogenizes institutions (p.14) and endorse the recommendation that schools play a significant role in developing their own statements of goals and assessment methods, but I was dismayed by the suggestion that the examining body might select from a pool of 100 to 500 test items. In the first place I doubt that most faculties can develop a pool of 500 items that goes beyond the most trivial levels of specific facts, and in the second place I believe that it is increasingly recognized that the most important kinds of thinking involve open ended problems in which there is no one right answer and no algorithm for arriving at the right answer. Thus it is unlikely that one can rely upon objective tests to assess the kinds of outcomes most colleges would like to believe they are achieving. This is not to say that some portions of the assessment could be scored objectively, but there is a danger in putting heaviest weight upon those outcomes which are easiest to assess.

In Appendix A I would endorse again the aim of improving education. There is a danger that one gets so wrapped up in the technicalities of testing that one loses sight of education. In this regard, I believe that one should try to develop tests that provide useful information for students as well as faculty and administrators. This probably means aiming for meaningful scores on component skills rather than global measures. It also implies that attention be given to the way in which results are communicated to students and the public as well as to higher education personnel.

I also repeat my endorsement of faculty involvement, but I would add two codicils, one to include experts in evaluation conceived broadly, as well as experts in test construction, particularly those skilled in non-objective testing, the second to revise the assessment measures periodically, not only to keep them up to date but also to insure the continued cooperation of the faculty as faculty groups change through new appointments, changing responsibilities, retirements, etc.

This reminds of a general observation. It seems to me that you want faculty and students to feel committed to the assessment as fair, useful,

and valid. Yet you don't want teachers to teach to the test or students to study for the test. Since the tests used can never be more than a sample of the whole domain of educational outcomes, it loses its validity if all that a student learns is that that is tested.

I liked the paragraph that begins toward the top of page 21, and I hate to see it relegated to an appendix. I'd like these sentiments to come both at the beginning and conclusion as well as in the Executive Summary.

I hope these comments are helpful. Please don't take them as criticisms for I think you've done a very good job--in fact amazingly good within the time you were allowed.

You asked me to specify a fee, and I judge that $250 would be appropriate for the time involved. I estimate that my expenses for telephone etc. would not exceed $15. My Social Security # is 362-22-7068.

I enjoyed and learned from your report. Best wishes for the New Year!

Bill McKeachie
4660 W. Joy Rd.
Dexter, MI 48130

NOTE TO HECBOARD:

Professor Wilbert McKeachie is Associate Director of the Center for Research on Learning and Teaching at the University of Michigan.

SCHOOL OF EDUCATION
STANFORD UNIVERSITY, STANFORD, CALIFORNIA 94305

N. L. GAGE
MARGARET JACKS PROFESSOR OF EDUCATION
AND, BY COURTESY, PROFESSOR OF PSYCHOLOGY

December 29, 1966

Dr. Fred E. Fiedler
Organizational Research Group
Department of Psychology
NI-25
University of Washington
Seattle, Washington 98195

Dear Fred,

In addition to the comments I gave you over the phone on December 28 and 29, I offer the following remarks concerning the draft of your report, "The Evaluation of Institutions of Higher Education in the State of Washington," to assist the Higher Education Coordinating Board of the State of Washington.

1. Overall, the report impresses me as valid, thoughtful, and comprehensive, especially in view of the short period in which it had to be prepared.

2. In some places, the report seems to deal with goals and objectives other than those of "assessing institutional performance of the State's four-year college programs." I refer particularly to the places where you refer to the goals of proper student placement and guidance.

3. You never refer explicitly to the important distinction between norm-referenced and criterion-referenced evaluation. The former kind of evaluation makes comparisons with other students or departments or institutions. The latter requires comparisons with necessarily subjective and absolute standards. In many ways, the criterion-referenced kind of evaluation will be necessary. The judgment of competent persons will be inescapable, in large part because norm-referenced comparisons will be invalid, for the reasons you give at various points. For example, if 70% of the graduates of the University of Washington indicate that they are happy, that percentage will need to be interpretated in and of itself rather than compared with percentages obtained from other universities, other states, and so on.

4. To a large degree, the Higher Education Coordinating Board is, or should be, concerned with the degree to which the State's universities are serving the goal of social equity. Higher education traditionally provides an avenue toward upward mobility. To what extent does the State's system of higher education serve this purpose? To what extent does the system serve students from low-income families? The answers are to be found, in part, in statistics concerning the numbers of applicants, admissions, retentions, and graduations of students from low-income families, ethnic minorities, and other traditionally disadvantaged groups. Your proposal should be amended to provide for adequate investigation and evaluation of phenomena pertinent to such questions. Indeed, almost all of the statistical analyses you mention

in the draft might well be analyzed, insofar as possible, according to different income-level and ethnic-group status. Furthermore, of course, what I have just said applies to the differences to be found in all of your statistics between men and women if justifiable concerns about gender equity are to be served.

5. It might help to clarify some of your major points if you were to indicate that the most justifiable basis for evaluating the impact of higher education would require a true experiment--one in which students were divided at random into college-going and noncollege-going groups. Such a randomized pretest-posttest control-group design would be sufficiently valid to provide a truly defensible evaluation of the impact of the State's universities. It is also, of course, ethically impossible. But it would serve to clarify for your readers the kinds of limitations as to validity that must characterize the feasible kinds of evaluations.

6. A point at random: The $40.00 fee for your respondents seems high; $10.00 or $20.00 might well do the trick. Perhaps some pilot testing of the necessary and desirable fee would be worthwhile.

7. After describing the difficulties and invalidities of the "value-added" approach to evaluation, you go on to deal with the ways in which it could be applied. Your convincing discussion of its limitations makes the treatment of its uses and advantages seem beside the point.

8. Nowhere did you discuss, as I recall, another major difficulty in attributing posttest status, alumni beliefs, alumni activities and employment, and the like, to the causal influences on competence of going to college. Rather, the influences of going to college may result merely from the college's serving as a credential-awarder or gate-keeper. Just as knowing Latin was associated with higher status and achievement in nineteenth-century England not because it caused greater social and economic competence but because it served as a credential that permitted young men to pass through the gates toward higher social and political status, so college-going may be associated with high scoial status and achievement merely because employers want people with college diplomas and not because people with college diplomas are necessarily more capable of doing a job or serving a community. How you can get around this possible misinterpretation of the effects of college-going (i.e., attributing higher economic achievement to higher competence resulting from college going rather than to higher social acceptability resulting from having a college diploma) is a problem that your present draft does not confront. It is a problem that will especially beset your interpretation of your alumni surveys but not, as I see it, your measures of student achievement of the objectives of general education or the objectives of their majors. The difference is that the achievement measures are aimed at maximum performance, which is not influenced by mere gate-keeping and credentialling functions of higher education. But beliefs, attitudes, community activities, job status, income, and the like, are easily seen to be influenced by the mere credentialling and gate-keeping functions of higher education. The latter functions (gate-keeping and credentialling) can of course take valid forms, as when only high-achieving students are admitted to medical school. But they can also take invalid forms as when brokerages insist on employing only college graduates, even though what was learned in college courses may be unnecessary or irrelevant to the performance of a job in the brokerage; the college diploma here serves as an index of social acceptability, the "school tie," rather than of intellectual and social competencies produced by the college and needed by the employer.

I shall not try to specify the many points, propositions, facts, distinctions, cautions, and observations that I found myself admiring in your report. For lack of time and in whole awareness of your January 6 deadline, I want to get this letter into your hands as soon as possible. In my opinion you have done an admirable job in the time available. I only hope that some of my observations prove to be useful and useable.

Sincerely yours,

Nate

N. L. Gage

P.S. Because you may find my compulsive copy-editing useful, I'm returning herewith the copy of the report. N

NLG:pb

APPENDIX G

REFERENCE LIST

ACT National Center for the Advancement of Educational Practices. ( ) Assessing student competence and persistence: the best case for general education.

Adelman, A. (1986) Assessment in American higher education, Washington, DC: Office of Educational Reseach and Improvement-US Department of Education.

Assessing the Outcomes of Higher Education. (1985) Proceedings of the ETS Annual Invitational Conference. 25 October 1986.

Astin, A.W. (1970) The methodology of research on college impact: part one. Sociology of Education, 43 (1970), 223-254.

Astin, A.A. (1977) Four critical years. SF,CA:Jossey-Bass.

---------- (1985) Achieving educational excellence. SF,CA: Jossey-Bass.

Banta, T. ( ) Assessing outcomes: the real value-added is in the process. TN:U of Tennessee.

Boyer, E. (1986) (Unknown title) The Chronicle of Higher Education, 15 October 1986, 41.

Bover, E.M. & Hollister, W.G. Behavioral science frontiers in education. NYC: Wiley.

Carroll, B. (1969) Job satisfaction. Ithica, NY: Cornell U.

Cook, J.D., Hepworth, S.J., Wall, T.D., & Warr, P.B. (1981) The experience of work: a compendium and review of 249 measures and their use. SF,CA: Academic Press.

Evangelauf, J. (1986) National reports on undergraduate education spur changes at 1 in 3 colleges, survey finds. The Chronicle of Higher Education, 30 July 1986, 20-21.

Ewell, P. (Ed.)(1985) Assessing educational outcomes: new directions for institutional research. SF,CA: Jossey-Bass.

Evell, P. ( ) The self-regarding institution: information for excellence. ?:National Center for Higher Educational Management Systems.

Evell, P. & Jones, D. ( ) The costs of assessment. Boulder, Colorado: Wiche.

Gruneberg, M.M. (1979) Understanding job satisfaction. London: Macmillan.

In pursuit of degrees with integrity: a value-added approach of undergraduate assessment. ( ) MS:NE Missouri State U.

Jaschik, S. (1986) Get more involved in college reform, state legislators told. The Chronicle of Higher Education, 30 July 1986.

Jaschik, S. (1986) Governors call on all colleges and universities to develop comprehensive programs to measure student learning. The Chronicle of Higher Education, 3 September 1986, 70-78.

Pace, C.R. & Stern, G.G. An approach to the measurement of psychological characteristics of college environments. Journal of Educational Psychology, 49(5), 269-277.

Portigal, A.H. (Ed.) (1973) Measuring the quality of working life. Ottawa: Nev Research Initiatives Research and Development Program.

Quinn, R.P. & deMandilovitch, M.S. (1977) Education and satisfaction: a questionable payoff. Ann Arbor, MI: Survey Research Center - U of Michigan.

Sevell, W.H. & Shah, V.P. (1968) Parents' education and children's educational aspirations and achievements. American Sociological Review, 33(2), 191-209.

Spaeth, G.L. & Greeley, A.M. (1970) Recent alumni and higher education. NY,NY; McGrav-Hill.

Stinzi, V.L. (1970) The relationships between the need for responsible work and job satisfaction of nev college graduates. Seattle, WA: UW.

Taylor, J.C. (1973) Concepts and problems in studies of the quality of working life. LA, CA:UCLA.

Taylor, J.C. (1973) The quality of working life: an annotated bibliography. LA, CA: UCLA.

Taubman, P. & Wales, T. ( ) Mental ability and higher educational attainment in the twentieth century. In...(to be provided).

Withey, S.B. (1975) A degree and what else? Correlates and consequences of a college education. NY: McGraw-Hill.

Wachtel, P. ( ) The returns to investment in higher education: another view. In...(to be provided).

Walsh, R.M. & Birkin, S.J. (Eds.) (1979) Job satisfaction and motivation: an annotated bibliography. Westport, CN: Greenwood.